The role of multisensory integration in the bottom-up and top-down control of attentional object selection

The role of multisensory integration in the bottom-up and top-down control of attentional object selection

Pawel Jerzy Matusz
Brain & Behaviour Lab
Department of Psychological Sciences
Birkbeck College - University of London

Submitted for the Degree of Doctor of Philosophy

March 2013

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own, and that this work has not been submitted for any other degree or professional qualification.

Pawel Jerzy Matusz

19th March 2013

Publication note

Experiment 1 and Experiment 5 reported in Chapter 2 and Chapter 3 of this thesis, respectively, have been published in Psychonomic Bulletin & Review. Experiments 8, 9, and 11 reported in Chapter 5 are under review for publication in the journal Psychophysiology.

Abstract

Selective spatial attention and multisensory integration have been traditionally considered as separate domains in psychology and cognitive neuroscience. However, theoretical and methodological advancements in the last two decades have paved the way for studying different types of interactions between spatial attention and multisensory integration. In the present thesis, two types of such interactions are investigated.

In the first part of the thesis, the role of audiovisual synchrony as a source of bottom-up bias in visual selection was investigated. In six out of seven experiments, a variant of the spatial cueing paradigm was used to compare attentional capture by visual and audiovisual distractors. In another experiment, single-frame search arrays were presented to investigate whether multisensory integration can bias spatial selection via salience-based mechanisms. Behavioural and electrophysiological results demonstrated that the ability of visual objects to capture attention was enhanced when they were accompanied by non-informative auditory signals. They also showed evidence for the bottom-up nature of these audiovisual enhancements of attentional capture by revealing that these enhancements occurred irrespective of the task-relevance of visual objects.

In the second part of this thesis, four experiments are reported that investigated the spatial selection of audiovisual relative to visual objects and the guidance of their selection by bimodal object templates. Behavioural and ERP results demonstrated that the ability of task-irrelevant target-matching visual objects to capture attention was reduced during search for audiovisual as compared to purely visual targets, suggesting that bimodal search is guided by integrated audiovisual templates. However, the observation that unimodal target-matching visual events retained some ability to capture attention indicates that bimodal search is controlled to some extent by modality-specific representations of task-relevant information.

In summary, the present thesis has contributed to our knowledge of how attention is controlled in real-life environments by demonstrating that spatial selective attention can be biased towards bimodal objects via salience-driven as well as goal-based mechanisms.

Table of Contents

List of Figures

To my partner, Raphael, without whom none of this would have been possible,

and to my mom, who always tirelessly believed in me

Acknowledgements

I am deeply grateful to my supervisor Professor Martin Eimer. Without his guidance, patience and down-to-earth approach this thesis would not have been possible. Thank you for teaching me never to lose sight of 'the bigger picture' in my research.

I would also like to thank Dr Monika Kiss, Dr Angela Gosling, and, especially, Dr Anna Grubert, for discussions, as well as their help and advice. Special thanks for the lab and technical support of Sue Nicholas, without whom successful completion my research would be much more difficult. I would also like to thank Dr Antoine Spiteri, Dr Przemek Tomalski, Dr Rachel Wu, and Dr Silvia Dalvit, who provided me with invaluable help and advice at various stages of my Ph. D. studies.

I am most deeply indebted to Magda Bednarz - thank you for your continuous support throughout the past three and a half years. I would also like to thank Jakub Traczyk and Agata Sobków - research is always so much fun with you, guys.

This work was in part supported by the Research Studentship from Birkbeck College.

Chapter 1. General Introduction

1.1. Brief introduction to multisensory integration and visual attention

Effective cognitive functioning and behaviour is determined by the ability to encode and interpret events that are potentially or directly important to our goals. A network of sensory systems, where each is specialised in processing a particular type of input, jointly increases the chance of correctly identifying events that are most relevant in different circumstances. However, separate mechanisms are necessary to efficiently select meaningful events from the multi-modal stream of stimuli registered by the senses at every moment. Traditionally, researchers argued that selective processing is executed only by attention – a set of mechanisms operating at different stages of information processing through which processing of the sensory input crucial to current behavioural goals of an individual is enhanced, while the irrelevant input is filtered out. The majority of studies on attention have focused on the visual modality (Corbetta & Shulman, 2002; Desimone & Duncan, 1995; Morris, Öhman, & Dolan, 1999). However, research in the past two decades has shown that the brain has a strong tendency to integrate information presented to different senses, on the basis of close temporal/spatial proximity or semantic congruence. There is growing evidence that such integrated stimuli often have a competitive advantage for processing resources over unimodal events (Koelewijn, Bronkhorst, & Theeuwes, 2010). This process of combining inputs across modalities, known as 'multisensory integration' (Stein & Stanford, 2008), was recently shown to take place at different stages of information processing (Driver & Noesselt, 2008), suggesting different potential loci for mutual interactions with visual attention during the process of stimulus selection. Critical advancements have also been made recently in understanding how visuo-spatial attention is controlled in real-life environments, through investigations into how salience-based and goal-driven mechanisms interact in the selection of objects in space (Desimone & Duncan, 1995; Wolfe, 2007). In spite of the progress that has been made in both areas, systematic research into the role of multisensory integration in the control of orienting of spatial attention in environments where multiple objects are simultaneously present is still largely lacking.

The aim of this thesis was to investigate two major forms in which selection of objects in multi-stimulus contexts can be biased by multisensory integration. The first form

of selection bias driven by multisensory integration includes situations in which a combination of signals from different modalities occurs automatically and prior to the stage at which information is selected. Fusion of different sources of stimulation should result in the creation of an emergent, salient multimodal event, and this event should have an enhanced ability to attract visuo-spatial attention to its location in an involuntary manner, compared with visual stimuli in the multi-stimulus context. The investigation of factors modulating such multisensory enhancements of attentional capture will greatly improve our understanding of the cross-modal mechanisms of bottom-up control of object selection, an issue that has been largely overlooked in the existing literature. The second way in which multisensory integration can bias selection of objects in space is observed in situations in which target stimuli are defined by a known combination of features in different modalities. Here, attention is directed in a top-down fashion towards multimodally defined target stimuli, and the allocation of attention in space is controlled by a multimodal task set. However, up to this point it has not been investigated whether spatial selection can be biased towards bimodal objects in a top-down fashion, i.e., on the basis of integrated multimodal object templates, which are defined as a conjunction of specific features from different modalities. If such highly selective multisensory top-down guidance is possible, then searching for objects which are defined as conjunctions of visual and auditory features (e.g., a red bar paired with a high-pitch tone) should result in suppressed selection of stimuli which match only one of these features (e.g., all red bars presented without high-pitch tones). Investigating whether search for objects defined across modalities can be guided by integrated object templates will provide a much more ecologically valid picture of attentional control, by revealing how attention is controlled in contexts in which stimuli are defined by combinations of features from different modalities.

This chapter begins with a brief introduction to multisensory integration and to visual attention (Section 1.1), followed by a description of the neural substrates of both cognitive processes (Section 1.2). The aim of these sections is to summarise the current state of knowledge on the functional interactions between multisensory integration and visual attention, as well as to highlight important gaps in this knowledge (Section 1.3). Next, event-related potentials are described as an invaluable method in addressing the open issues concerning the interaction of multisensory integration and selective attention (Section 1.4). Finally, the two fundamental forms of this interaction, that are the focus of this thesis, are outlined (Section 1.5).

1.1.1. Multisensory integration

1.1.1.1. Behavioural advantages of multisensory integration

Until recently, the dominating view on sensory processing was that initial perceptual analysis occurs in respective sensory pathways in a modular fashion, while the integration of its outcomes takes place only at later stages, in the so-called 'association' or 'polysensory' brain areas. A consequence of this view has been that research in the areas of psychology and neuroscience was carried out on each modality in isolation, an approach that was not only understandable, given the scarceness of knowledge on cortical organisation at the time, but also extremely valuable in that it allowed the fundamental principles of perceptual processing within each modality to be established (Evans & Whitfield, 1964; Hubel & Wiesel, 1962; Mountcastle, 1957). Meanwhile, throughout the 1980s and 1990s, major advancements in neuroimaging, including such novel methods as functional magnetic resonance imaging (fMRI), positron emission tomography (PET) and magnetoencephalography (MEG), meant that researchers could now study the functioning of the whole brain in living organisms. These theoretical and technological developments together have led several researchers to conclude that a full understanding of perceptual processing can only be achieved by finding out how the processing of stimuli in different senses affects each other (see Calvert & Thesen, 2004).

The ability to combine information across modalities has evolved due to the adaptive advantages it brings (Calvert, 2001). Two fundamental ways in which multisensory integration can benefit unimodal processing (see Alais, Newell, & Mamassian, 2010, for a review) are by: (1) disambiguating or enriching the information that is conveyed unimodally (i.e., enabling the cognitive system to make a critical categorisation about the event or object of interest); and (2) providing the same information about a stimulus of interest that is readily available in the primary modality (i.e., enabling the cognitive system to benefit from this information redundancy).

The first benefit of multisensory integration for information processing is well exemplified by the coherence of perception of objects in multimodal environments. Typically, modalities provide complementary information about the same property of an external event. Features from dimensions, such as intensity, shape, texture, rate, rhythmic structure, but also spatial and temporal location, provide analogous information regardless of the modality ('intermodal invariance'; Lewkowicz, 2000). However, in contexts where the input from one modality is insufficient to correctly recognise an object, information from

another modality can enrich the perceptual representation by providing the missing information. For example, Newell, Ernst, Tjan, and Bülthoff (2001) showed that when the back of a three-dimensional object was not visible to the participant, haptic exploration of the object was critical for recognition. In other instances, information from the secondary modality can disambiguate the input provided by the primary modality. The 'stream/bounce illusion' is a good example of a situation in which input to one modality alone supports different perceptual interpretations. More specifically, when two disks move towards each other from opposite sides of a screen, it is difficult to judge whether they stream through or bounce off each other when coinciding in the centre. Sekuler, Sekuler, and Lau (1997) showed that presentation of a sound concurrently with the movement of two disks coinciding greatly aids perceptual discrimination by strongly supporting the 'bouncing' interpretation.

As regards the second benefit, redundancy of information conveyed by different modalities benefits perception and behaviour as it reduces the overall amount of information that needs to be processed for an object to be perceived accurately and acted upon appropriately (Lewkowicz, 2000). Early studies on cross-modal interactions (Hershenson, 1962; Stein, Meredith, Huneycutt, & McDade, 1988) showed that a target can be responded to faster when accompanied by a stimulus in another modality, even when the latter bears no additional information other than the temporal coincidence. Various explanations were offered over the years for this so-called 'cross-modal redundant target effect' (Bernstein, Clark, & Edelstein, 1969). It was interpreted as reflecting the sum of activities elicited separately by each stimulus (Miller, 1982), or mere statistical facilitation (Raab, 1962). However, recent behavioural and ERP studies (Fort, Delpuech, Pernier, & Giard, 2002a; Gondan, Goetze, & Greenlee, 2010; Gondan, Niederhaus, Rösler, & Röder, 2005; Molholm, Ritter, Murray, Javitt, Schroeder, & Foxe, 2002; see also Grubert, Krummenacher, & Eimer, 2011, for similar effects for within-modal pairings) demonstrated that facilitation of processing based on redundancy of signals from different modalities is driven by the actual *integration* of activations elicited by signals, which can take place at the stage of perceptual analysis, and which results in more effective processing at later stages (such as the execution of a motor response). Notably, loci for this type of multisensory integration seem to be present at various stages of information processing (Fort, Delpuech, Pernier, & Giard, 2002b; Fournier & Eriksen, 1990; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; Miller, 1991).

Current interest in multisensory processing (Alais et al., 2010) has been triggered by a series of seminal studies conducted by Barry Stein and colleagues into the neuro-cognitive

mechanisms underlying the benefits of spatiotemporal redundancy in orienting to peripheral stimuli (for a review, see Stein, 1998). Their behavioural findings are easier to understand from the point of view of the underlying neural mechanisms. By using single-neuron recording in deep layers of cat's the superior colliculus (SC), a structure implicated in overt and covert attention shifts (i.e., shifts accompanied and not accompanied by eye movements, respectively), Meredith and Stein (1986) revealed that combining redundant spatial and temporal information from different modalities leads to a pattern of neural enhancements that are not observable for redundant pairings of unimodal signals, i.e., the firing rates of neurons in the SC in response to the bimodal stimulus were higher than the most effective unimodal stimulus.

Critically, these enhancements were visible only if certain principles concerning the relations between the unimodal stimuli were fulfilled. Two rules concerned, respectively, the redundancy of spatial and temporal information conveyed by two signals. Stein and colleagues argued that this is likely due to the fact that presentation of two signals from roughly the same region in space indicates a single location of potential importance, and temporal co-occurrence (see also Meredith, Nemitz, & Stein, 1987) suggests a single stimulus as their source. Notably, the largest enhancements of neural processing ('superadditive' responses; for more details, see Section 1.2.1) were observed when one of the inputs was of low intensity. This effect embodied the third rule, the 'inverse effectiveness rule', i.e., the strongest benefits of multisensory integration should be found for unimodal stimuli that are weakly effective and could be missed when presented separately. Meredith and Stein (1986) regarded the neural enhancements triggered by spatiotemporally redundant signals as evidence of the generation of a new unified multimodal stimulus from a combination of two unimodal stimuli. Subsequently, this neural marker and the three principles described above became a framework that was used in research on the forms of multisensory integration that occur in the cortex (see Calvert & Thesen, 2004).

On the basis of a clear causal relationship between the activity of the SC and orienting behaviour (through descending projections to brainstem and spinal cord; see Meredith & Stein, 1986), Stein and colleagues (1988) investigated whether the behavioural benefits in detection of bimodal stimuli in peripheral space follow the same rules as found for multisensory enhancements in the SC. Four cats were trained to orient to and approach (for a food reward) the location of a dim and difficult-to-detect flash of an LED light that could be presented in one of several possible locations arranged equidistantly in a semicircle. In conditions where the visual target was accompanied by a spatiotemporally

aligned tone (neutral in respect to reinforcement), the percentage of correct detections of the LED light was larger than the sum of correct detections of the unimodal stimuli presented in isolation. In contrast, when the same 'neutral' tone was spatially misaligned with the LED flash, the bimodal target was detected significantly less frequently when compared with the visual target (and triggered a 'response depression' of multisensory neurons in the SC; see Stein, 1998). The neural and behavioural findings of Stein and colleagues provided the first evidence that, if specific conditions are fulfilled, multisensory integration can strengthen the representation of a unimodal object, which in turn can enhance orienting behaviour.

The evidence that human observers can detect a degraded peripheral visual stimulus more accurately when it is paired with a spatiotemporally aligned non-visual signal has been provided only recently. Frassinetti, Bolognini, and Ladavas (2002) conducted a behavioural study using an experimental setup similar to the one employed by Stein et al. (1988). Participants were asked to indicate the detection of a faint flash at one of several possible peripheral locations by pressing a button. The flash was presented alone or accompanied by a task-irrelevant tone. Critically, the spatial and temporal proximity of the tone in respect to the flash was systematically modulated. In order to decide whether effects of multisensory integration affected the perceptual or the later, response-related, stages of processing of the visual stimulus accompanied by a spatiotemporally aligned tone, modulations of the parameters of the signal detection theory (Tanner & Swets, 1954) were investigated. In psychophysics, the signal detection theory is typically employed to quantify the ability of a detecting system to discriminate information-bearing energy patterns (i.e., stimulus) from random energy patterns (i.e., noise; background stimuli and random activity of the nervous system of the observer), and to assess how different factors can affect the detection threshold applied in this process.

Frassinetti et al. (2002) argued that if concurrent presentation of a tone renders the visual stimulus more distinguishable from its background, an increase of the perceptual sensitivity parameter *d'* (or 'd prime'), indexing better discrimination of signal from background noise, should be observed at the location of the target. In turn, if the bimodal stimulation increases merely the participants' tendency to press the button, this response bias change should be visible as an increase of the decision criterion parameter, β (or 'beta'). The results showed that *d'* was increased on tone-present versus tone-absent trials, indicative of multisensory integration enhancing detection of peripheral visual objects by strengthening their perceptual representation. However, when the two signals were misaligned, i.e., the tone was presented at a different location or 500 ms apart relative to the light flash, this facilitation was eliminated, suggesting that in human observers the benefits

in the detection of faint peripheral stimuli associated with audiovisual integration follow the same spatial and temporal principles as those found in a cat's SC (Stein et al., 1988).

1.1.1.2. Cross-modal transfer of spatial and temporal information

Faster and more accurate identification of perceptual input are those consequences of multisensory integration that provide individuals with an advantage during interactions with the environment. However, transfer of information across senses is so pervading that its influence is largely uncontrollable even in situations in which incongruent information is provided by different modalities, which results in various forms of perceptual illusions. To demonstrate the robustness of cross-modal interactions, the following section will briefly discuss the illusions that are most relevant to the topic of this thesis.

In the context of sensory conflicts arising due to incongruent spatial information, ventriloquism is the most widely known effect. In this illusion we have an impression that speech produced by a puppeteer is being produced by the puppet. It is suggested that such a 'capture' of the auditory stimulus to the location of a visual event that is temporally coincident has the benefit of improved perception under noisy circumstances, i.e., temporally concurrent cross-modal signals will be bound together in the location of the visual stimulus (Sumby & Pollack, 1954). The current literature suggests that ventriloquism is largely automatic in that it can be observed even with simple auditory and visual stimuli (e.g., tones and flashes, respectively) lacking any semantic content: In a series of experiments, Slutsky and Recanzone (2001) revealed that the illusion takes place as long as two events are in close temporal and spatial proximity (cf., Meredith & Stein, 1986). Further evidence for the automaticity of the illusion was provided by findings showing that ventriloquism occurs even when attention is focused elsewhere in space, or when participants are directly instructed to ignore the visually presented information when trying to localise the tone (Bertelson & Radeau, 1981; Vroomen, Bertelson, & de Gelder, 2001a, 2001b). The principle of 'visual capture' that underlies ventriloquism seems to explain also the 'audiovisual apparent motion' effect (Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002). This illusion describes the fact that observers who are instructed to indicate the direction of apparent motion created by a series of separate auditory stimuli often report that the sound is coming from concurrently presented visual events, and not from the actual location of the auditory stream. This evidence strongly suggests that the tendency of the brain to combine information across modalities is so profound that it can

lead to the location of one event in space being perceptually 'captured' to the location of a concurrent but task-irrelevant visual distractor.

Just as widespread are illusions based on 'auditory capture', where the perception of the point of time at which an event of interest happens is biased towards the onset time of an auditory stimulus presented in close temporal proximity. This effect, also known as 'temporal ventriloquism', was investigated by Morein-Zamir, Soto-Faraco, and Kingstone (2003). Participants assessed which of two LED lights appeared first ('temporal order judgement') in a context where each light was accompanied by an irrelevant tone. If the two sounds were presented so that one preceded the first flash and the other followed the second flash, participant judgements were more accurate than when each flash was synchronised with a sound. In other words, judging which flash was presented first was easier in the former compared to latter condition, as if the accompanying sounds putatively increased the interval between the two flashes. In a follow-up study, Vroomen and Keetels (2006) demonstrated that temporal order judgements are facilitated by neighbouring versus simultaneous tones even in cases where there are large spatial discrepancies between these tones (e.g., when they are presented from the opposite side of fixation). An important conclusion that can be drawn from the research on temporal ventriloquism is that not all cross-modal interactions require spatial correspondence (cf., Frassinetti et al., 2002).

Numerous illusions have demonstrated that the auditory system is superior to the visual system in respect to parsing a stream of stimulation into separate events (Welch & Warren, 1980). One of the more striking examples of how temporal information from audition influences visual perception is 'auditory driving', described first by Shipley (1964). In this illusion, participants judge the rate of a flickering light and initially irrelevant auditory clicks are presented synchronously with the light flashes (e.g., at a frequency of 10 cycles per second). Shipley (1964) showed that the changing rate of the auditory clicks could bias the flash rate judgement downward, to 7 cycles per second, as well as upward, to 22 cycles per second. Importantly, this involuntary transfer of temporal information does not occur when the task-relevant and irrelevant modality is reversed, suggesting that biasing the perception of rhythm through perceptual organisation is specific to the auditory system. A more contemporary form of this effect was recently studied by Shams, Kamitani, and Shimojo (2000). In the 'double flash illusion', there is a very strong tendency to perceive two flashes instead of one in contexts where a brief single flash is accompanied by two short sounds. Importantly, this effect occurs as long as the time window between the respective flash and sound is not larger than 100 ms (Shams, Kamitani, & Shimojo, 2002). Cross-modal interactions typically show this temporal constraint, with the likelihood of

multisensory integration decreasing sharply beyond this time interval (e.g., van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008a, Experiment 5; Vroomen & de Gelder, 2000).

Overall, the cross-modal perceptual illusions described above suggest that there may be a certain specialisation between modalities, whereby the visual system is dominant in the processing of spatial information and the auditory system dominates in the processing of temporal information. This idea was formalised by Welch and Warren (1980) as the 'modality appropriateness hypothesis', which represents the first attempt to explain the plethora of cross-modal perceptual illusions being discovered at the time. Welch and Warren argued that the superior spatial resolution of vision and superior temporal resolution of audition complement each other in creating coherent perception. However, their model has been revealed to be more of a useful simplification than an overarching principle that applies to all cross-modal phenomena in perception. For example, spatial ventriloquism was found also for audio-tactile pairings (Guest, Catmur, Lloyd, & Spence, 2001), thus arguing against a modality-specific nature of this phenomenon. Crucially, Alais and Burr (2004) showed that in cases where the visual stimulus is degraded (i.e., blurred Gaussian blobs), and thus poorly localised, tone location dominates localisation judgement, i.e., the visual dominance in spatial ventriloquism is reversed.

The growing evidence indicating that the cognitive system is much more flexible than the rigid dichotomy argued by Welch and Warren (1980) was well accommodated by an alternative account, rooted in the Bayesian probability theory. The 'optimal integration' (Ernst & Banks, 2002) or 'maximum likelihood estimation' (MLE; Ernst & Bülthoff, 2004) model proposes that the cognitive system combines signals from different modalities with the purpose of providing a joint estimate of a property of an external object (e.g., its location) that has the highest probability of being accurate. To yield such a statistically optimal estimate, the MLE sums up two signals, but in this process both signals are weighted by their reliability (i.e., reliable signals receive high weights and unreliable signals receive low weights). The less reliable component will not drive the bimodal estimate, but it will still contribute to it to a certain extent, effectively increasing its reliability. Because the combination of two inputs is more reliable than either signal separately, the outcome of the combination can be regarded as 'optimal'. Numerous studies have demonstrated that humans typically combine information from different modalities in this statistically optimal fashion (for a review, see Ernst & Bülthoff, 2004), and that they can do so in an automatic, attention-independent manner (e.g., Helbig & Ernst, 2008). Notably, the natural domination of vision in spatial tasks, and of audition in temporal tasks, is also consistent with this

model, thus rendering it a more flexible and quantitative alternative to the earlier 'modality appropriateness' model.

Statistically optimal multisensory integration has the power to explain a broad spectrum of phenomena, ranging from diverse perceptual illusions arising from interaction of spatial and temporal information across senses, to facilitation of perceptual and response-related processing of unimodal stimuli by temporally coincident task-irrelevant signals from other modalities. Most importantly, close temporal proximity between signals from different modalities might often suffice for these signals to be integrated in an effortless, attention-independent fashion into a single multimodal object.

1.1.1.3. Cross-modal interactions affecting basic perceptual judgements

There is already a substantial body of evidence suggesting that multisensory integration is more pervasive than previously thought, and that it affects even very basic judgements that were traditionally regarded as totally sensory-specific (see Driver & Noesselt, 2008, for a review). A pioneering study in this area by Stein, London, Wilkinson, and Price (1996) revealed that a concurrent but fully task-irrelevant sound can increase the perceived brightness of a light flash. Similarly to the study of Frassinetti et al. (2002), the largest enhancements were observed for visual stimuli with the lowest intensities, near the perceptual threshold. Critically, Stein et al. (1996) demonstrated that not all cross-modal interactions follow the three principles governing multisensory enhancements in the SC and orienting behaviour. They showed that, for concurrent visual and auditory stimuli presented at peripheral locations, spatial alignment between two signals was critical for the perceived intensity of the flash to be enhanced, as indexed by subjective reports of brightness. However, when the stimuli were presented so that the flashes appeared in a central location where subjects were fixating, perceptual benefits were observed despite large spatial disparities between the two stimuli. These findings were the first to suggest that multisensory integration can affect perception merely through an increase of the salience of the primary stimulus (see Gillmeister & Eimer, 2007, for similar results involving audiotactile pairings).

However, as Stein and colleagues (1996) used a subjective measure of perception, it was unclear whether the reported tone-induced enhancements were driven by a genuine increase in the brightness of flashes that were accompanied by sounds, or simply by a stronger tendency to categorise audiovisual stimuli as brighter than visual stimuli (Odgaard,

Arieh, & Marks, 2003). Importantly, Stein et al.'s (1996) findings were subsequently replicated and extended. Using a contrast detection task, Lippert, Logothetis, and Kayser (2007) demonstrated that pairing low-contrast Gabor patches with task-irrelevant and spatially diffuse tones leads to increases in the *d'* parameter measured in response to such stimuli, thus suggesting their enhanced distinctiveness from the background. Audiovisual synchrony was also shown to facilitate discriminative processes. In a study by Noesselt, Bergmann, Hake, Heinze, and Fendrich (2008), participants indicated which of two vertically aligned dotted circles presented simultaneously on the screen disappeared for a brief moment. On trials in which this offset was paired with a spatially diffuse tone, performance (measured by RTs and error rates) was improved in comparison to tone-absent trials. Thus, studies using more objective measures of perception than Stein et al. (1996) provided strong evidence in support of the notion that temporal coincidence of stimuli in one modality can increase the distinctiveness of centrally presented objects in another modality. These salience-based effects in perception, where a signal in another modality increases the quality of perceptual representation of objects in the task-relevant modality, are likely to be driven by a neural mechanism in which signals from different modalities converge during the initial feedforward processing sweep (Lamme & Roelfsema, 2000) from lower- to higher-level sensory brain areas (see Foxe & Schroeder, 2005, for a review).

To summarise, there is a wide range of perceptual and behavioural effects, from puzzling perceptual illusions to the facilitation of object recognition and orientation in space, as well as very basic, 'sensory-specific' judgements. Different forms of multisensory integration seem to underlie these cross-modal phenomena, supported by different neural substrates (as will be shown in Section 1.2.1), and following different principles that are necessary for their occurrence. An important conclusion that arises from the plethora of cross-modal interactions in perception is that the cognitive system has a pervasive tendency to reduce the absolute amount of information to be processed at any point in time, for effective functioning, and it will frequently achieve this by combining signals that are likely to belong to the same object or event.

1.1.2. Visual attention

1.1.2.1. Pioneering theories of visual attention: space- and object-based attention

The second most fundamental way in which the cognitive system selectively processes objects and events important to current behavioural goals is selective attention. William James (1890) was the first to argue that attention acts as an 'internal spotlight' within which specific stimuli are processed in great detail (see also Broadbent, 1982). Early studies in visual attention suggested that the primary medium of selectivity is location, and Posner (1980) conducted pioneering research into how selective spatial attention facilitates perception.

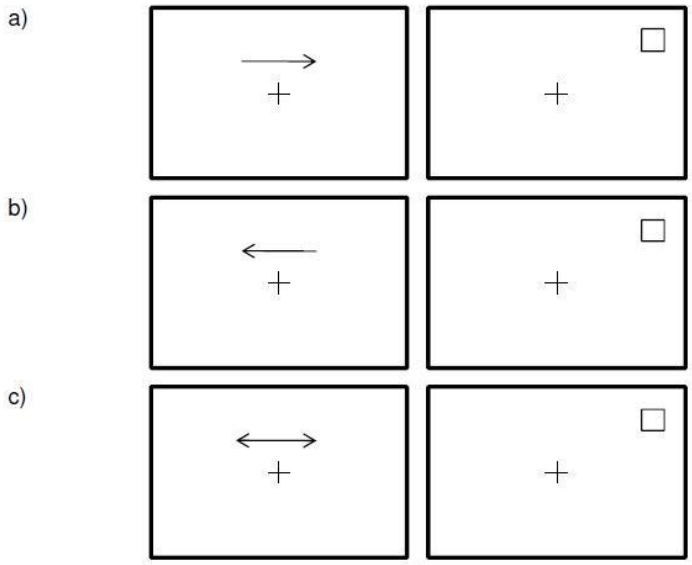

Figure 1.1. *A version of spatial cueing paradigm experiment developed by Posner (1980), with a) valid, b) invalid, and c) neutral trials. The target in this example is represented by a square presented in the panels on the right.*

Posner (1980) developed an experimental paradigm in which participants were covertly shifting their attention; participants responded to the onset of a flash occurring in one of several possible locations in the display, while their eyes were fixated on a point in the middle of a screen. Targets were preceded by spatial cues, 'valid' on a majority of the trials (i.e., indicating the likely location of the target; Figure 1.1, panel a) and 'invalid' on a minority of the trials (i.e., the target would appear on the side opposite to the cue; Figure

1.1, panel b). Locations could be indicated by a central cue (i.e., a centrally presented arrow pointing to the left or right side of the display) or a peripheral cue (i.e., a brief illumination of the box in which the target would appear). A neutral cue, which provided no information about the likely target location (e.g., a centrally presented cross, Figure 1.1, panel c), was also occasionally presented.

Posner (1980) demonstrated that directing attention in advance to the location of an upcoming target resulted in a faster analysis of perceptual features of the target. Valid trials showed attentional benefits (shorter RTs than on neutral trials), while invalid trials showed attentional costs (longer RTs than on neutral trials). Critically, peripheral cues showed costs and benefits even in contexts where they correctly indicated the target location on a minority of trials. These results motivated Posner (1980) to propose the existence of two fundamental attentional systems: (1) an 'endogenous system', controlling attention shifts in a voluntary, or 'top-down', manner on the basis of observer's goals and expectations; and (2) an 'exogenous system', responsible for involuntary, or 'bottom-up', shifts of attention to the location of a salient peripheral event. Subsequently, another fundamental difference between endogenous and exogenous cues was revealed: The effects of the latter on orienting behaviour are short-lived (Müller & Rabbitt, 1989). If the time interval between the cue and target is long (e.g., larger than 500 ms), exogenous cues lead to responses to targets presented at their location, which are even longer than on uncued trials. This effect is typically known as 'inhibition of return' (Posner & Cohen, 1984), and is interpreted as reflecting a natural tendency of attention to avoid regions of space that were recently attended to.

The 'spotlight hypothesis' (Broadbent, 1982), according to which the focus of attention can be moved around the visual field, and enhances processing of stimuli that fall inside that focus, can account for the results from Posner's cueing paradigm, as well as for the findings from other paradigms, e.g., the flanker task. Eriksen and Eriksen (1974) showed in this task that identification of a target letter is delayed by presence of flanker distractors that are associated with a different response, but only in contexts where the targets and distractors appear in close spatial proximity (i.e., 1° of visual angle); no interference is found for distractors presented at larger distances from the target. These findings suggested that the attentional 'beam' has a fixed size within which all stimuli are automatically selected. Additionally, Eriksen and Yeh (1985) showed that it is difficult to direct attention to more than one location at a time. When participants were cued to two opposite locations in which the target was equally likely to appear, their performance was facilitated only in response to targets presented at the location where they actually appeared, but not at the

opposite, equally probable location. These results were taken as evidence that the attentional 'spotlight' cannot be effectively 'split' simultaneously between two different locations (but see Awh & Pashler, 2000, for evidence for multiple spotlights). However, a growing body of evidence, reported at the same time as the studies described above, suggested an alternative perspective: The nature of attention is more similar to a 'zoom lens', where the resolution of the attentional focus can be adjusted as a function of task demands (Downing & Pinker, 1985; Eriksen & Murphy, 1987). More specifically, the 'spotlight' has the ability to be both diffuse, i.e., enhance (but only weakly) perceptual processing of all stimuli present in the visual field, and to be narrowed down, enabling in-depth analysis of a single event (Eriksen & St. James, 1986).

The 'spotlight' hypothesis provided an important novel way of thinking about visual attention as space-based in nature, where selection of task-relevant information occurs through the location of objects in space. However, identification of objects in space, rather than mere selection of locations in space, seems to be the most fundamental function of visual attention in the external environment, which is predominantly composed of multiple objects. This 'object-based' nature of attention was first suggested by the work of Gestalt psychologists who showed that our mind has a natural tendency to group stimuli in the visual field together. More recently, converging behavioural, neuropsychological and neuroimaging evidence was provided for this account (Duncan, 1984; Fink, Dolan, Halligan, Marshall, & Frith, 1997; O'Craven, Downing, & Kanwisher, 1999). In a seminal study, Duncan (1984) demonstrated that in contexts where two separate attributes need to be identified (e.g., a rectangle with a gap on the left vs. right side, and left-ward vs. right-ward tilt of a vertical line), errors are less frequent in cases where these attributes belong to a single object than if they belong to two different objects. Notably, the object-based model of attention can also explain the results taken to support the 'spotlight' hypothesis. For example, the response interference between the target and its flanker distractors observed by Eriksen and Eriksen (1974) could have been triggered by target and flanker distractors being processed as a single object due to their close spatial proximity. What is important, rather than being mutually exclusive, the space- and object-based forms of attention seem to interact, jointly affecting target detection (e.g., Egly, Driver, & Rafal, 1994). Thus, the research discussed in the present section should be regarded as evidence for flexibility of the attentional system, with different forms of selection used interchangeably or jointly for effective perception and behaviour.

1.1.2.2. Searching for specific features in visual space: FIT

While the space- and object-based selection accounts describe many principles that underlie orienting behaviour, they cannot fully explain why searching for objects in space can often dramatically differ in its difficulty. For example, if we are searching for our favourite blue jumper in the bedroom, we will most probably have a sensation that it literally 'pops out' in front of our eyes when there are no other blue items in the bedroom at the time (i.e., distractors). This effect can be explained by an account that proposes that object features are the primary form of attentional selection. The two main theories supporting this account are 'feature integration theory' (FIT; Treisman & Gelade, 1980; Treisman & Sato, 1990) and the 'guided search' model (Wolfe, Cave, & Franzel, 1989; see Wolfe, 2007, for the latest version of the model). These models argue that there is a limited number of object categories, or 'dimensions', such as colour, orientation, curvature, size or motion, which can be used to characterise objects. More specifically, a jumper can be described through several dimensions, such as colour, shape and size, but also by using specific features, such as red or blue (colour) and big or small (size).

To systematically investigate why, in certain circumstances, searching for a target in space does not seem to be affected by the presence of surrounding distractors, Treisman and Gelade (1980) developed the 'visual search' paradigm. In this task participants are typically presented with an array of stimuli in which they need to detect the presence of a particular target by pressing a button.

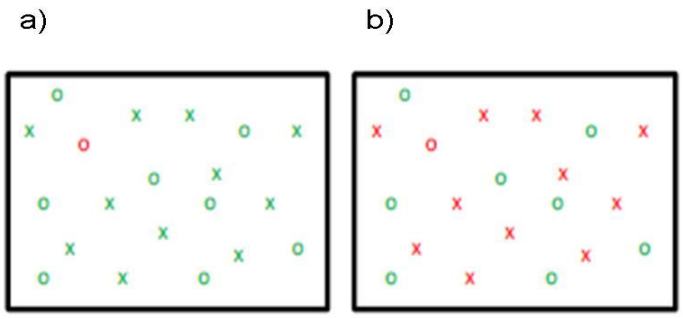

Figure 1.2. *Examples of visual search display with target defined as red O. The panels depict search arrays with a) feature target, and b) feature-conjunction target, respectively.*

Treisman and Gelade (1980) showed that there is a difference between searching for an object characterised by a single feature ('feature search') and searching for an object characterised by two features ('conjunction search'). In contexts where participants were searching for a red O among green X's and O's (see Figure 1.2., panel a), search was relatively fast and not affected by the number of surrounding distractors ('flat search function', i.e., search times do not increase even if the number of distractors does). In contrast, searching for a red O among red X's and green O's (see Figure 1.2, panel b) triggered overall slower RTs, which increased proportionally with the number of distractors ('steep search function').

Treisman and Gelade (1980) proposed the 'feature integration theory' to explain these divergent findings. According to this model, there is a set of spatiotopically organised maps hardwired in the brain, which code the presence of elementary attributes in the visual field in a parallel and automatic manner. If a target is characterised by a feature discontinuity (a feature singleton), it will be detected effortlessly through the activation it elicits in a relevant feature map. At successive stages of visual processing, the location of this activity is coded by a master map, leading to an automatic shift of attention in this direction (the target 'pop-out' effect). However, if a task requires detection of a target characterised by a combination of two features, these features need to be joined correctly at the same location in the master map. Because attention is applied sequentially to locations indicated by the master map, search for a feature conjunction is necessarily serial, as attention can only be deployed to a single location in the visual field at a time. Hence, Treisman and Gelade (1980) argued, spatial attention is also important in correct object identification, as during conjunction search it acts as 'glue' that joins separately processed features into a unified object representation.

1.1.2.3. Critique of the FIT model

The FIT model was supported by evidence from beyond the visual search paradigm, such as physiological data indicating that the early analysis of different features takes place in different brain areas (e.g., Maunsell & Newsome, 1987) and connectionist models of vision suggesting that serial processing is an optimal strategy for correctly integrating different features belonging to an object (e.g., Feldman, 1985). While the theory was very influential as the first systematic attempt to delineate the nature of the search process, numerous criticisms arose over the years, which could not have been easily accommodated by the model (for a comprehensive review, see Quinlan, 2003). The most important limitation of

the FIT model concerns the proposed simple dichotomy between parallel, pre-attentive feature search and serial, attentive conjunction search, which dichotomy could not explain the great variability of search slopes that were reported by a rapidly growing body of research. Various studies demonstrated that flat and steep search slopes depict only two extremes of what is a continuum of search functions, with some instances of feature search producing steep RT search slopes, and some instances of conjunction search showing search functions independent of the number of distractors (see Wolfe, 1998, for a review). The evidence that search efficiency is described by a continuum argues strongly against the FIT as a model accurately describing the visual search process.

According to another criticism of FIT, it is not the nature of the stimuli, but the relations between stimuli in the visual field that determine whether search is efficient or inefficient. In a series of studies, Duncan and Humphreys (1989, 1992) showed that the similarity between target and distractors and the similarity among distractors are crucial determinants of the efficiency of the search process. Duncan and Humphreys (1989) created the 'search surface' in order to depict how these two dimensions affect RTs (see Figure 1.3). When a target is highly dissimilar from distractors, search slopes are flat (line AC), even if it is defined as a conjunction of features. Similarly, when distractors are very similar to each other, search is efficient (line AB). In general, the more similar target and distractors are and the less similar distractors are among each other, the steeper the search slope and the more difficult the search becomes (towards point D).

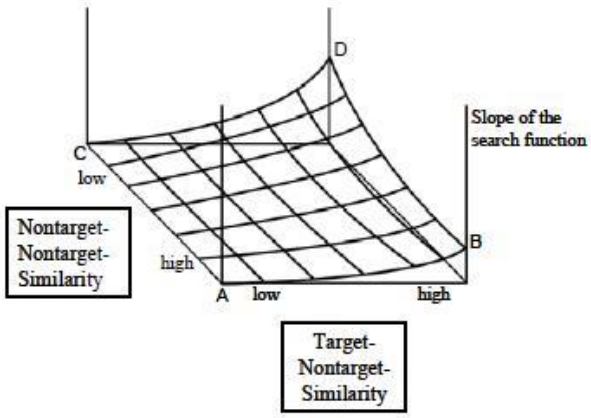

Figure 1.3. *Schematic illustration of the search surface based on target-distractor (dis)similarities as proposed by the attentional engagement theory, adapted from Duncan and Humphreys (1989).*

The FIT model was developed further by Wolfe (1994, 2007) in his guided search (GS) model. According to this theory, the initial analysis of dimension-based feature representations leads to computation of the saliency values, which indicate the presence of features at specific locations in the visual field. The model also proposed that the second stage encompasses a topographically organised 'priority map' ('overall saliency map' in the earlier versions of GS) which, in contrast to FIT, does not receive any direct inputs from the feature representations, but integrates (i.e., summates) the activations from different dimension-based saliency maps. As a result, focal attention that operates on the priority map is directed to the location of the highest salience values (as opposed to being directed to the location of a feature, as proposed by FIT) and subsequently grants to features present at this location access to the higher-level cognitive processes concerned with object recognition and response preparation. If the currently analysed stimulus matches the definition of the target (i.e., the 'target/ search template'), a response is executed; if it does not match the template, attention is allocated to the location exhibiting the second highest level of activation and this process is repeated until the target is found.

In another contrast to FIT, the GS model proposed that both the pre-attentive and the attentional stages are involved in detection of simple features and feature conjunctions. Having incorporated both bottom-up, stimulus-driven (in line with the early versions of FIT), and top-down, goal-driven mechanisms of selection control, the GS model has the power to explain visual search effects that range from 'pop-out' search to the continuum of search efficiency. Namely, when searching for a target that can be defined merely as 'the odd stimulus in the display' (i.e., a feature singleton), the automatic, stimulus-driven guidance of attention to the location of the highest salience activation can fully explain why such a target is instantly detected. In contrast, during search for a conjunction of features (e.g., a blue vertical line), knowledge-driven mechanisms will trigger activation in all locations containing blue or vertical stimuli, and the target location will be indicated as the location of the highest activation. In other words, in contexts where a large number of distractors are sharing one of the features of the target, the search times are longer due to a higher level of noise in the priority map. Despite the high explanatory power of GS, some visual search effects, such as shorter RTs in cases where the target-defining dimension is constant across trials (e.g., Found & Müller, 1996; Müller & Krummenacher, 2006), cannot be easily reconciled with this model.

1.1.2.4. The role of top-down and bottom-up factors in control of visual selection

While the distinction between exogenous, stimulus-driven and endogenous, goal-driven visual selection seems to be conceptually clear-cut, the relative importance of bottom-up and top-down factors in the control of attentional shifts has become a point of a major debate in the attentional research in the past twenty years (Beck & Kastner, 2009; Desimone & Duncan, 1995).

According to some researchers (Theeuwes, Atchley, & Kramer, 2000; Theeuwes, 2010), local differences in contrast among neighbouring stimuli in the visual field always determine which location is going to be selected first, irrespective of whether the identity of the current target is known or not. Theeuwes (1991) created a derivative of the visual search task, also known as the 'additional singleton paradigm', in which he demonstrated that search for a known target (a shape singleton, e.g., unique diamond) among uniformly shaped distractors (e.g., circles) will be delayed if an irrelevant but salient feature-singleton distractor (e.g., a red circle) is present in the search array. Theeuwes (1991) regarded these results as evidence that attention will always be initially captured by the most salient item in the visual field. However, these findings contrasted sharply with the results of Folk and colleagues (Folk, Remington, & Johnston, 1992) who showed that attentional selection of items in the visual field is determined by current behavioural goals (i.e., the 'task set'). Folk and colleagues designed the spatial cueing paradigm, in which search display was preceded on each trial by a cue array. In this paradigm, the ability of stimuli to involuntarily attract attention is measured through spatial cueing effects, i.e., shorter RTs on trials in which the target location is cued compared to trials in which targets appear at other, uncued locations (cf., Theeuwes, 1991). Folk et al. (1992) demonstrated that salient feature singletons, e.g., red cues, triggered reliable cueing effects as long as the targets were defined by colour, but not when they were defined as abrupt onsets. These results led Folk and colleagues to propose the so-called 'task-set contingent involuntary orienting' hypothesis, according to which salient distractors will capture attention only in cases where they are task-relevant, i.e., if they share features with the target.

Various solutions were proposed to reconcile these apparently contradictory results. On the one hand, it was argued that delayed RTs in the additional singleton paradigm are caused by the additional time required by filtering mechanisms that are necessary to perform the search efficiently, but which are not connected to actual shifts of attention (Folk & Remington, 1998). On the other hand, Theeuwes et al. (2000) proposed that the most salient

item in the display will always be selected first but that attention can be rapidly disengaged when the item is identified as a nontarget, leading to no visible cueing effects, as shown by Folk and colleagues (1992, 1994). The most plausible theoretical solution was offered by Bacon and Egeth (Bacon & Egeth, 1994), who argued that two different 'search modes' were encouraged by demands of the additional singleton paradigm and the spatial cueing task. In the Theeuwes (1991) study, the target could be effectively detected merely by monitoring the display for an 'odd one out' ('singleton-detection mode', SDM). The costs of attentional capture by the singleton distractor were likely to be minimal as the search array remained presented until response, thus allowing participants to reorient to the feature singleton target.

Bacon and Egeth (1994) argued that the low demands that SDM places on the cognitive system render it a 'default' search strategy. In contrast, in the study of Folk et al. (1992), the cue preceded the target by 150 ms and the target appeared only for 50 ms, with no time to inspect the target in case of attentional capture by a salient distractor. In this context, the low cognitive demands associated with SDM would be outweighed by substantial costs in search performance, thus forcing observers to search for a specific feature in a pre-defined dimension, e.g., all red targets ('feature-search mode', FSM). In support of this account, Bacon and Egeth (1994, Experiment 2) showed absence of RT costs during search for a target diamond in the presence of a colour singleton, in a version of the Theeuwes' (1991) search task, in which other uniquely shaped distractors (e.g., triangles, squares) were present in the array, which rendered SDM ineffective. In other words, task demands, not local contrast differences, determined whether salient distractors captured attention, in line with the primary role of a goal-based mechanism in attentional control (but see Dalton & Lavie, 2004, for task-set independent capture of auditory attention by within-modal feature singletons). Recently, this hierarchy between top-down and bottom-up factors in the control of visual attention shifts has been further supported by studies which manipulated these factors in a systematic manner (Eimer, Kiss, Press, & Sauter, 2009; Eimer & Kiss, 2008, 2010; Lamy & Egeth, 2003; Lamy, Leber, & Egeth, 2004; Lien, Ruthruff, Goodin, & Remington, 2008): Converging behavioural and electrophysiological evidence demonstrated that the role of stimulus-driven factors in the control of visual selection is at best indirect or secondary (see Section 1.4.3.3, for details).

The plethora of findings indicating prevalence of top-down mechanisms in the control of visual selection was integrated in the 'biased competition model' of visual attention (BCM; Desimone & Duncan, 1995; Duncan, Humphreys, & Ward, 1997). The novel perspective on visual selective attention that this model proposed was based on an

older idea that processing in most visual brain systems is based on competition. On the basis of neurophysiological and behavioural findings (e.g., Bundesen, Shibuya, & Larsen, 1985; Chelazzi, Miller, Duncan, & Desimone, 1993; Duncan, 1984), Desimone and Duncan (1995) proposed that in multi-stimulus contexts, stimuli compete with each other for control over the receptive fields of neurons that lead to representation of a specific stimulus in the visual cortex. Competition is already reflected by the overall lower number of spikes triggered in a neuron in response to displays with two stimuli, compared to each stimulus presented alone, suggesting the presence of inhibitory interactions between two stimuli. Furthermore, competition occurs at any level of cortical processing at which several stimuli fall within the receptive field of a single neuron, i.e., competitive interactions are stronger at higher levels of cortical hierarchy (e.g., the inferotemporal cortex and posterior parietal cortex), as neurons at these stages have the largest receptive fields, often encompassing both left and right hemifields.

The critical tenet of BCM is that competition is typically modulated by factors concerned with the observer's goals, but also with bottom-up salience of stimuli in the external environment. In other words, there are numerous neural mechanisms (see Section 1.2.2) that favour stimuli that match current task requirements (i.e., flexible change of medium of selection between objects, locations and features), but competition is also biased towards stimuli that are novel or highly different from their surroundings. In multi-stimulus contexts in which one or both mechanisms give a competitive advantage to one of the objects (e.g., in 'pop-out' displays in which only targets of pre-defined identity are presented), the competitive interactions are resolved in favour of this object, which is reflected by firing rates approximating the sum of firing rates to the stimuli presented alone. BCM also proposes that the resolution of competition between multiple stimuli is frequently integrated across different levels of cortical hierarchy, therefore, an object in favour of which the competition was resolved at the level of perceptual processing, should also be the one that 'wins' the competition at the stage at which stimuli compete to be encoded into short-term memory or control shifts of visuo-spatial attention (covert or overt). The presence of biased competition in multi-stimulus arrays and the integration of competition resolution across neural systems have been substantiated by fMRI studies (Beck & Kastner, 2005, 2007; Kastner et al., 1998; Reynolds & Desimone, 2003) and recent ERP studies (e.g., Kiss, Grubert, & Eimer, 2013). The critical novel contribution of BCM into the understanding of the nature of selective attention is that it defines selective attention not as a 'spotlight' that moves across distinctive locations in empty space to facilitate perception or bind separate features, but as an emergent state of resolved competition across neural networks involved

in perceptual processing and behavioural control (see Beck & Kastner, 2009, for a related view on the nature of attention within the BCM).

In summary, early research on visual attention demonstrated that attention facilitates perception and behaviour through an interaction of two separate but interlinked attentional systems: the endogenous attention system that enables us to focus on stimuli relevant to our current behavioural goals, and the exogenous attention system that diverts our attention to potentially important events in the periphery. More recent findings have highlighted the flexibility of goal-based control mechanisms in guiding attention towards goal-related information on the basis of locations (both in space and time), objects or features. These discoveries have jointly paved the way for investigations into how attention operates in real-life environments, where multiple objects compete for selection at any point in time, and many of these objects are defined across different modalities.

1.2 Neural substrates of multisensory integration and visual attention

1.2.1. Neural substrates of multisensory integration

Early neuroanatomical studies using various species of non-human animal subjects and employing diverse methods of measuring localised brain activity led to the view, which prevailed in the 1960s and 1970s, that multisensory integration is limited to the association cortices that are related to late stages of information processing. In line with this approach, an area would be classified as a 'convergence zone' if it was shown to contain neurons receiving afferent inputs from multiple modalities. On the basis of this definition, 'convergence sites' were discovered in various cortical and subcortical areas, with new areas successively added to the list. Cortical areas currently regarded as multisensory are the superior temporal sulcus, posterior parietal cortex, frontal regions including the premotor, prefrontal, orbito-frontal and anterior cingulate cortex, as well as the insula and hippocampus (Bruce, Desimone, & Gross, 1981; Grunewald, Linden, & Andersen, 1999; Jones & Powell, 1970, see Ghazanfar & Schroeder, 2006, for a review). Multisensory subcortical areas include the claustrum, tectum, the pulvinar nuclei of the thalamus, and the SC (Fries, 1985; Mesulam & Mufson, 1984; Pearson, Brodal, Gatter, & Powell, 1982, see Calvert, 2001, for a review).

1.2.1.1. Traditional multisensory convergence zones

As discussed in Section 1.1.1.1, single-neuron recording studies conducted by Stein and colleagues in cats' SC led them to the discovery of neuronal mechanisms underlying multisensory integration and the principles that govern this process at the level of a single neuron (see Stein & Stanford, 2008).

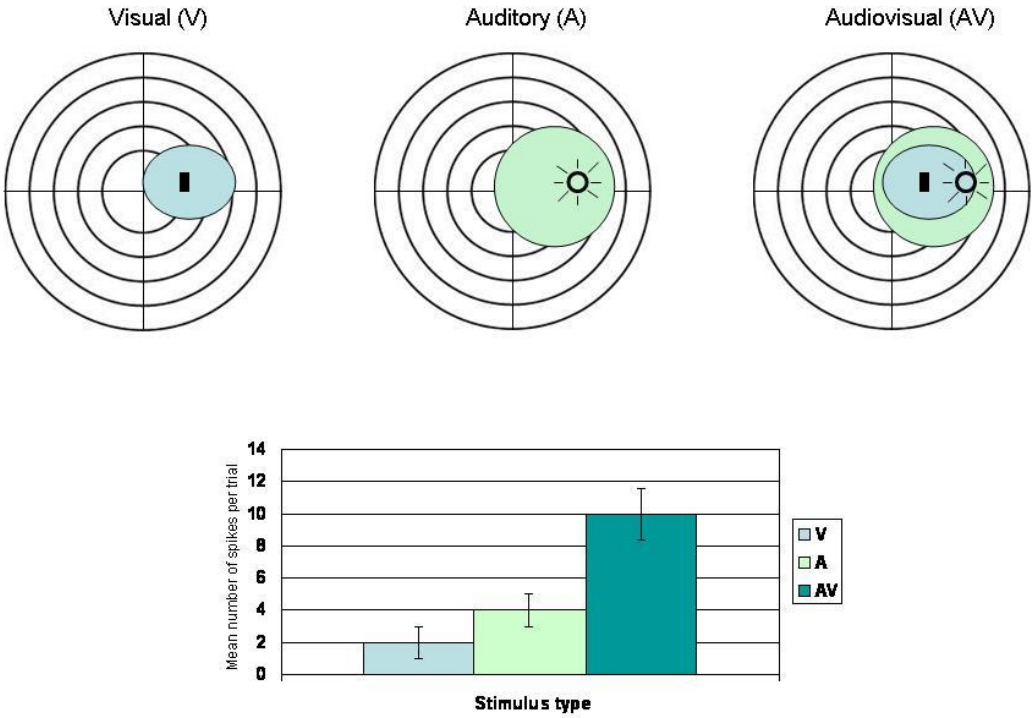

Figure 1.4. *Schematic illustration of multisensory integration in a single neuron located in deep layers of SC, adapted from Meredith and Stein (1986). The visual and auditory receptive fields of the neuron and locations of stimuli within visual and auditor space are shown. A weakly effective visual stimulus and an auditory stimulus are integrated, what produces multisensory enhancement of neural responses. In the example shown in the figure the response to bimodal stimulus exceeded the sum of unimodal responses.*

Deep layers of the cat's SC were shown to contain neurons that respond to inputs from different senses, such as vision, audition and touch. More specifically, each neuron contains different receptive fields, one for each modality, which are aligned in close 'spatial register' (i.e., they overlap with each other, see Figure 1.4, left panel). Such organisation

creates a functional multisensory map of external space, in which the emphasis is put on the location of a stimulus, rather than the modality, and which enables the SC to trigger an accurate orienting response. Meredith and Stein (1986) showed that two signals in close spatiotemporal proximity trigger a number of action potentials ('spikes') that is larger than the number of spikes triggered by the stronger signal (i.e., 'multisensory enhancement' of the neural response). Thus, rather than merely responding to signals from different modalities, SC neurons are involved in an integrative process, where bimodal events is processed differently than a sum of unimodal signals.

In some cases, the neural responses measured by Stein and colleagues to bimodal stimuli were exceeding the sum of spikes elicited by two signals presented in isolation (i.e., 'superadditivity' of the neural response, see Figure 1.4, right panel). The enhancement of the neural response can be maximal when the individual stimuli are only weakly effective and could go undetected if presented alone, but this principle of 'inverted effectiveness' was subsequently shown to be far from ubiquitous (e.g., Stanford, Quessy, & Stein, 2005). Some researchers even argued for it to be merely an artefact driven by the chosen analysis method (Holmes, 2007). A notable observation made by Meredith and Stein (1986) was that if two simultaneous signals are presented from disparate locations, a 'response depression' is observed, whose function is to inhibit responses to signals that do not originate from the same external event. In comparison, the temporal window in which two signals can be integrated was shown to be quite broad, which seems to allow the SC to accommodate for large temporal differences between stimuli caused by transduction times, neural latencies and speed of light and sound (Meredith et al., 1987).

Since 1980s, numerous brain areas in monkey and human brain were described as showing some form of multisensory integration. This led Ghazanfar and Schroeder (2006) to propose a provocative argument that the whole neocortex, to a smaller or larger extent, might be capable of combining signals from different modalities. Notably, the principles discovered by Stein and colleagues were adapted in a somewhat dogmatic manner as a framework for investigating cross-modal interactions in the cortical areas. Crucially, many researchers in the area regarded the superadditivity of neural response as an index of multisensory integration, and did not consider the possibility that the phenomenon could be a specialised computational solution that evolved to support effective localisation of genuine targets among spurious activity in the external environment (see Alais et al., 2010, for a review). In contrast to SC, the function of cortex is to accurately represent objects and semantic information, which suggests that the congruence of input provided by different modalities should be the critical factor for multisensory integration. In this context,

congruence should be understood as object- or action-related appropriateness of meanings that are carried by the two different signals, e.g., a sound of forging, but not of walking, would be congruent with an image of a hammer.

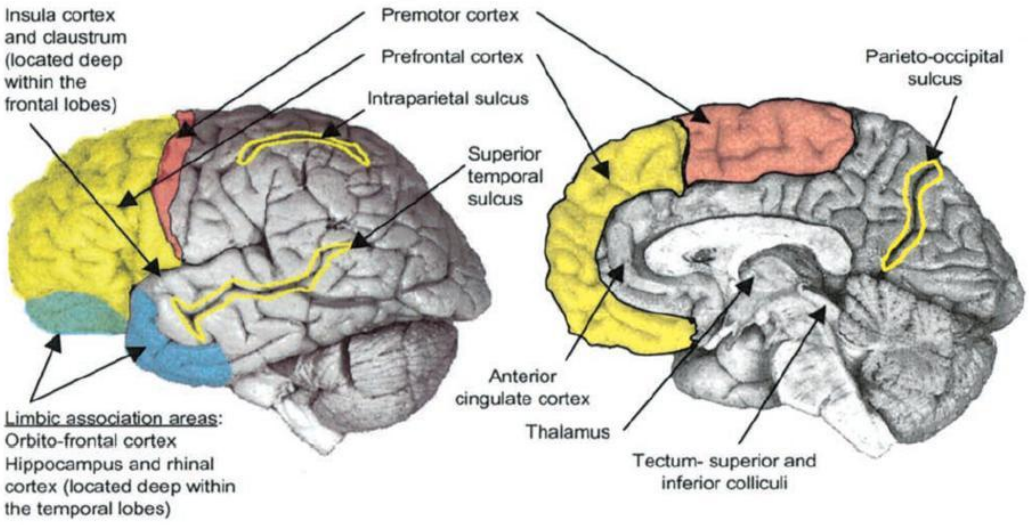

Figure 1.5. *Areas of putative heteromodal zones in the human brain shown on lateral (left) and medial (right) views, based on analogy from primate data, human neuropsychological evidence and human functional imaging studies. Calvert, G. A., Crossmodal processing in human brain: Insights from functional neuroimaging studies, Cerebral Cortex, (2001), 11, 1110-1123 by permission of Oxford University Press.*

In primates, one of the cortical areas often reported as containing populations of multisensory neurons is the intraparietal cortex, which is a part of a larger network located in the posterior parietal cortex (PPC; Graziano, 2001) that is responsible for multisensory-guided movements in space. The intraparietal cortex (see Figure 1.5), composed of several subregions (lateral intraparietal, LIP; medial intraparietal, MIP; ventral intraparietal, VIP), is involved in goal-directed attention shifts. This heterogeneous area fulfils its function by integrating cross-modal signals into spatial maps that have the ability to re-map to a common eye-centred coordinate frame (but see Alais et al., 2010). This common frame of

reference provides the necessary basis for computing a vector for gaze towards audio-visual stimuli in LIP, as well as a vector for goal-directed limb movement in MIP. Notably, in spite of the fact that VIP contains neurons with SC-like spatially overlapping receptive fields, enhancement-type and depression-type neural responses to spatiotemporally aligned stimuli seem to be equally likely in this area, as demonstrated by one of the very few studies explicitly analysing multisensory integration in this structure (Avillac, Ben Hamed, & Duhamel, 2007). This suggests that the complexity involved in neural computations during multisensory integration is higher in neocortex than in the SC.

While within the intraparietal network cross-modal interactions might be critically dependent on action-related congruence between stimuli, in the superior temporal sulcus (STS; see Figure 1.5), a structure likely involved in processing of biologically relevant stimuli (e.g., speech or real-life objects), such relations seem to be determined by object-related congruence (i.e., an ecological match between signals) and familiarity with the object. Response enhancement (i.e., superadditivity), as well as responses depression (i.e., 'subadditivity' of neural response) was observed in the STS to concurrent naturalistic stimuli in single-neuron recording studies in macaques (Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004), and in neuroimaging experiments with human subjects (Calvert, Campbell, & Brammer, 2000; Foxe, Wylie, Martinez, Schroeder, Javitt, Guilfoyle, Ritter, & Murray, 2002; but see Hocking & Price, 2008).

Interestingly, a recent neuroimaging study by Werner and Noppeney (2010) showed that in a task involving categorisation of naturalistic objects (e.g., a dog) on trials, in which participants categorised audiovisual stimuli faster than unimodal stimuli, both STS and PPC were activated, thus suggesting their joint involvement in the maintenance of integration of higher-order object features. In comparison, prefrontal cortex (PFC) might be involved in the representation of novel but behaviourally meaningful cross-modal associations (e.g., Fuster, Bodner, and Kroger; 2000), which confirms its role as part of the endogenous attention network (e.g., Fockert, Rees, Frith, & Lavie, 2004). The research reviewed in this section highlights that multisensory integration in heteromodal or association cortical areas serves the function of coherent representation of objects or actions, in which semantic congruence plays a crucial role. Notably, in many cortical areas the functional role of multisensory integration is still not well understood, thus making unclear what its neural marker in each of these areas is (see Stein & Stanford, 2008).

1.2.1.2. Multisensory integration in low-level sensory cortices

In line with the psychophysical studies reviewed in Section 1.1.1, which demonstrated that cross-modal interactions can modulate basic sensory judgements, recent studies with monkey and human subjects showed that multisensory integration can also affect brain regions traditionally thought to be functionally independent and sensory specific. The majority of findings in this area was provided by anatomical tracing studies in animals, and revealed that different types of connections underlie these effects, ranging from feedforward convergence of inputs in specific low-level areas to feedback-type projections from higher-level, heteromodal areas (see Foxe & Schroeder, 2005; Kayser & Logothetis, 2007, for reviews). In primary visual cortices, such as V1 and V2, feedback projections from STS and PPC seem to be dominating, but sparse direct connections were also discovered between them and the primary auditory cortices, i.e., the auditory core (e.g., Falchier, Clavagnier, Barone, & Kennedy, 2002). It needs to be noted that some of these early-stage cross-modal connections might be based on indirect lateral connections via thalamus, an area long-known to be a 'relay' of sensory signals between various cortices (Cappe, Rouiller, & Barone, 2009; Hackett, de La Mothe, Ulbert, Karmos, Smiley, & Schroeder, 2007).

The function of such connections has not been fully understood, but the majority of the initial hypotheses point to the facilitation of unimodal processing by co-occurring signals in different modalities. For example, feedback-type connections from higher-level brain areas might support coherent multi-modal perception of the environment by disambiguating the feature analysis in low-level sensory areas (Kayser & Logothetis, 2007). This hypothesis is supported by large size of receptive fields in neurons in the heteromodal areas, which can provide large-scale scene information unavailable to the low-level neurons. In turn, the fact that the majority of auditory projections terminate in the areas of V1 and V2 that represent the peripheral visual field has led other researchers (e.g., Falchier et al., 2002) to argue that the role of cross-modal connections might lie in facilitating the detection of peripheral stimuli by enhancing their perceptual representation. Recently, Lakatos, Chen, O'Connell, Mills, and Schroeder (2007; see also Lakatos, Shah, Knuth, Karmos, & Schroeder, 2005) provided evidence for a likely physiological mechanism underlying integration of multisensory inputs in low-level cortices via direct feedforward and indirect lateral connections. Lakatos et al. (2007) analysed laminar profiles of synaptic activity (current source density, CSD) and multiunit activity (MUA) in primary auditory cortex of macaques. Their results demonstrated that signals from another modality facilitating unimodal processing by resetting the phase of ongoing oscillatory activity in low-level sensory areas

of the primary modality, which ensures that the crucial input arrives at the moment of the peak activity in the oscillatory cycle.

The research discussed in this section highlights that neural substrates for multisensory integration exist at very early stages of cortical processing, rather than being circumscribed to higher-level heteromodal cortices. Several theoretical models emerged in the recent years in an attempt to explain this plethora of cross-modal interactions. On one hand, the fact that some cross-modal interactions in low-level sensory cortices are supported by feedback connections from long-known multisensory convergence zones is in line with the traditional hierarchy of cortical processing. However, this account cannot fully explain the cross-modal connections supported by feedforward and lateral connections. On the other hand, the account proposed by Ghazanfar & Schroeder (2006), according to which the whole neocortex is to some degree multisensory, does not seem to hold true (but see Ghazanfar & Chandrasekaran, 2007; Kayser & Logothetis, 2007): Functional specialisation is a very basic rule of cortical organisation, and many brain regions reliably show a preference for signals from specific modalities (e.g., Macaluso & Driver, 2005). Thus, moderate versions of this model are more plausible. One of them argues that convergence zones are located at earlier stages of neural processing than traditionally thought, e.g., in various thalamic nuclei (Hackett et al., 2007), while another proposes entirely new converge zones, located in and around primary and secondary auditory cortex ('multisensory transitional zones', Wallace, Ramachandran, & Stein, 2004, see also Ghazanfar, Maier, Hoffman, & Logothetis, 2005; Lakatos et al., 2005). While more research is required in this area, the breadth of cross-modal interactions implies that all these models might be correct, as each describes a different interaction.

1.2.2. Neural substrates of visual attention

Selective attention facilitates perceptual processing of objects through the operation of two neural networks, which are responsible for instantiating and maintaining control ('sources' of attentional control) over the neural activity in the sensory visual areas ('sites' of influence of attentional control). The discovery of these networks through which visual attention controls selection in a top-down (endogenous attention) and bottom-up (exogenous attention) fashion was largely achieved by advances in heamodynamic neuroimaging methods, such as fMRI and PET. These methods rely on the assumption that there is a relation between neuronal activity induced by a task and the changes in blood metabolism, such as the level of blood oxygenation (also known as BOLD) in fMRI. Thus, inferences

about the underlying neuronal activity are drawn from changes in the cardiovascular system in a given brain area. In other words, heamodynamic neuroimaging methods are indirect measures of neural activity (cf., electroencephalography, Section 1.4.1).

To investigate which brain areas are recruited by the endogenous system, most studies used Posner's (1980) spatial cueing paradigm in which the likely location of an upcoming target is signalled by a central arrow cue (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Corbetta, Kincade, & Shulman, 2002). Measuring activity elicited by such cues can reveal the areas involved in top-down control of attention that are driven by knowledge and expectations (see Figure 1.6, panel a). The most reliable activations in response to such spatial cues were found bilaterally in intraparietal sulcus (IPS), superior parietal lobule (SPL), and the intersection between precentral sulcus (PrCes) and superior frontal sulcus (SFS; the intersection includes the putative human homologue of the frontal eye field, or FEF). Involvement of the dorsal frontoparietal network in the generation of preparatory states in attention, or top-down signals biasing neural processing (Desimone & Duncan, 1995), towards task-relevant features is further supported by findings from anatomical and physiological studies in animals (Colby, Duhamel, & Goldberg, 1993; Ungerleider & Desimone, 1986) and lesion studies in humans (e.g., Battelli, Cavanagh, Intriligator, Tramo, Hénaff, Michèl, Barton, 2001).

Unattended and low-frequency events also have the ability to control attention, and interrupt the task-at-hand (Arrington, Carr, Mayer, & Rao, 2000; Corbetta et al., 2000). The detection of such stimuli elicits responses in brain areas lying more ventrally compared to the dorsal network, most consistently in the temporal-parietal junction (TPJ) and in the ventral frontal cortex (vFC), which are both critical nods in the 'right ventral frontoparietal network' (see Figure 1.6, panel b). Neural activity in these areas is strongly lateralised, and is the strongest in response to targets presented at unattended locations. The 'right ventral frontoparietal' network is thought to serve the function of a 'circuit-breaker', i.e., it serves to re-orient attention to novel stimuli that can be potentially important (Corbetta & Shulman, 2002). Importantly, existing evidence suggests that activation of the right ventral frontoparietal network, in particular of the TPJ, is dependent on the behavioural relevance of a stimulus. Recent fMRI studies have revealed that this structure is activated by salient events only when they share features with the target (Serences, Shomstein, Leber, Yantis, & Egeth, 2005) or when the need for attentional control is low (e.g., the task involves only passive viewing; Downar, Crawley, Mikulis, & Davis, 2000). Numerous other findings (e.g., Eimer et al., 2009; Fockert et al., 2004) provide converging evidence that the exogenous

system of visual attention, which seems to be supported by the right ventral frontoparietal network, is strongly dependent on top-down attentional control settings (cf., Folk et al., 1992).

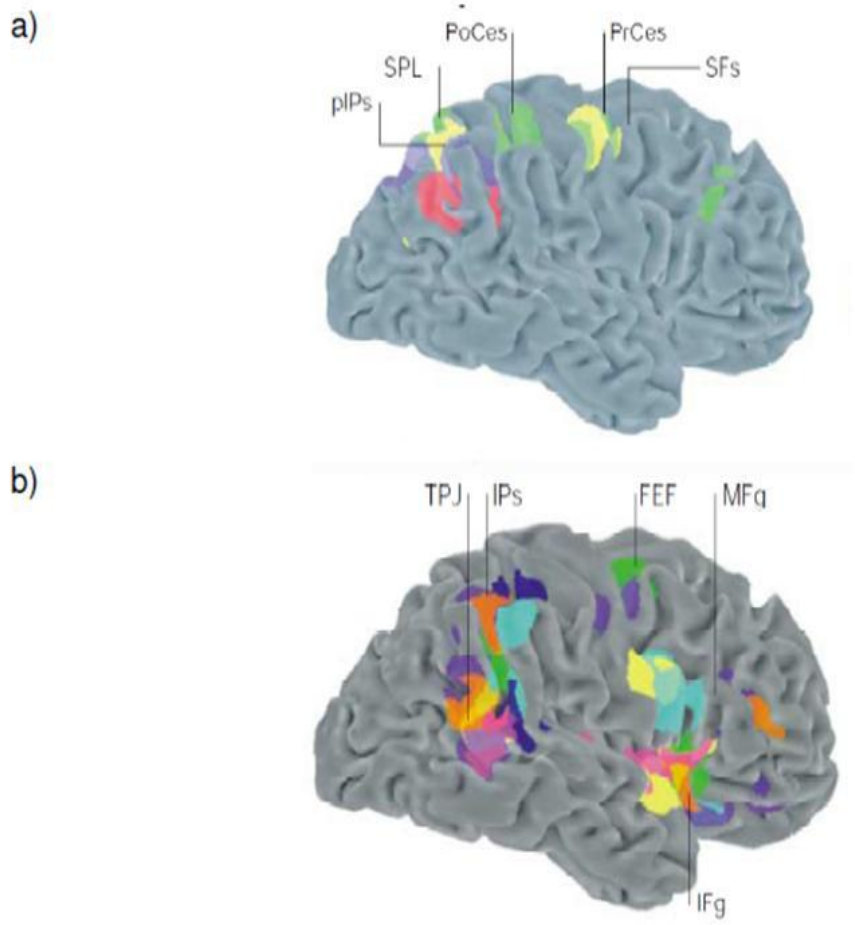

Figure 1.6. *Representation of the brain areas recruited by the a) dorsal and b) ventral fronto-parietal networks of attention. The dorsal fronto-parietal network includes the posterior intraparietal sulcus (pIPS), the superior parietal lobule (SPL), the postcentral sulcus (PoCes), the precentral sulcus (PrCes), and the superior frontal sulcus (SFs). The ventral fronto-parietal network includes the temporo-parietal junction (TPJ), the intraparietal sulcus (IPs), the frontal eye fields (FEF), the middle frontal gyrus (MFg), and the Inferior Frontal gyrus (IFg). Different colours represent varying results across studies. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience (Corbetta & Shulman, 2002), copyright 2013.*

There are two fundamental mechanisms by which top-down attentional control mechanisms facilitates perceptual processing of stimuli relevant to the task-at-hand. The first mechanism involves shifting attention to the object location, which enhances neural processing of this object in the cortices that code its location, as well as in the cortices coding features that define it. For example, O'Connor, Fukui, Pinsk, and Kastner (2002) carried out an fMRI study in which checkerboard stimuli were presented in the left or right hemifield, while subjects directed their attention to the stimulus location (attended condition) or away from it (unattended condition). When responses to attended and unattended stimuli were compared, the BOLD signal elicited by the former was significantly stronger than the signal elicited by the latter in various extrastriate areas in the hemifield contralateral to direction of attention, in line with the enhancing effect of shifting attention to a location on neural processing of an object in this location.

Importantly, top-down visual attention can also affect neural activity in the absence of actual stimulation: In a study by Kastner, Pinsk, De Weerd, Desimone, and Ungerleider (1999), a cue indicating the likely location of an upcoming target enhanced neural activity prior to the presentation of the target in all visual areas involved in representation of the attended location, which suggested that the baseline neural activity in task-relevant cortices can be increased by foreknowledge. This 'gain in sensory processing' mechanism is thought to reflect the top-down signals generated in the higher-level areas which enhance the sensitivity of neurons likely to be involved in neural processing of incoming input. Notably, neural activity in visual cortices that process the attributes of the target will be further enhanced during the actual presentation of the stimulus (Kastner et al., 1998).

It is important to note that there are cases, e.g. visual 'pop-out', in which visual selection is strongly affected by bottom-up signals. Such instances are well explained by the biased competition model (Desimone & Duncan, 1995; for more details, see Section 1.1.2.4.): The competition among stimuli in the visual field is likely to be resolved in favour of objects that are highly distinctive from their surroundings (Beck & Kastner, 2005; Reynolds & Desimone, 2003). This idea was supported by an fMRI study carried out by Beck and Kastner (2005), in which they presented four Gabor patches, and in different blocks one of these Gabor stimuli could differ from others in colour and orientation (a pop-out display) or all four differed from each other in both dimensions (a heterogeneous display). Results showed that the suppressive interactions, indicative of competition among simultaneous visual stimuli, disappear in cases where target objects appear in a pop-out context. Corroborating evidence for the notion that visual salience can bias competitive interactions was provided by studies using single-neuron recording (Nothdurft, Gallant, &

Van Essen, 1999) and computational modelling (Li, 1999) methods, all implicating area V1 as the possible source of bottom-up bias.

1.2.3. A comparison of the neural substrates of multisensory integration and visual attention

As described in the Sections 1.2.1 and 1.2.2, both multisensory integration and visuo-spatial attention are robust cognitive processes that recruit areas at multiple stages along of the structural hierarchy of neocortex. While there is a general consensus that attention operates through signals generated by two relatively independent neural networks in frontal and parietal cortex which affect posterior sensory areas, new multisensory convergence zones are constantly being discovered (see Driver & Noesselt, 2008). This suggests a complex network of interactions between visual attention and multisensory, with the nature of information processed and the task requirements likely to jointly influence the interplay. As discussed throughout Section 1.2.1, there are three possible models of how cross-modal signals are combined in the brain, i.e., early, parallel and late integration, each of which has different implications for interactions between multisensory integration and selective attention.

According to the late integration framework, signals from different modalities are first separately selected within modality-specific cortices and only then integrated in higher-level heteromodal cortical areas. The literature reviewed in the previous sections suggests that one of possible loci for multisensory integration at late stages of cortical processing might be IPS: On the one hand, this structure is a part of the dorsal frontoparietal network (Corbetta et al., 2000). On the other hand, IPS is involved in the maintenance of synthesis of visual and auditory attributes (Werner & Noppeney, 2010). Thus, IPS might be responsible for forms of multisensory integration which require separate selection of each unimodal signal within respective modality-specific cortices, e.g., the optimal cross-modal integration (Ernst & Banks, 2002; see Section 1.1.1.2) or search for objects defined by a conjunction of features from different modalities (see Chapter 5). However, more research is required to establish whether it is IPS that supports these forms of attention-dependent forms multisensory integration.

The early integration model, according to which signals from different modalities can interact at very early, pre-attentive stages of cortical processing, is supported by a recent but already substantial body of evidence (see Driver & Noesselt, 2008; Kayser & Logothetis, 2007, for reviews). Some of the instances of multisensory integration that take

place in low-level sensory cortices might occur through feedforward neural projections (Schroeder & Foxe, 2005). As it will be discussed in more detail in Section 1.4, EEG studies have provided strong evidence that cross-modal interactions can lead to enhancements in low-level sensory cortices at onset latencies that preclude explanations in terms of feedback projections from heteromodal cortical areas (e.g., Giard & Peronnet, 1999). Such perceptual illusions as the double-flash illusion (Shams, Kamitani, Thompson, & Shimojo, 2001) or ventriloquism (see Section 1.1.1.2, for details) underline the fact that multisensory integration can take place at pre-attentive stages of cortical processing, and possibly create a bottom-up bias towards bimodal objects during spatial selection at later stages.

The parallel integration framework argues that cross-modal signals can be integrated at different processing stages, and that the availability of attentional resources determines whether this takes place at early or later stages. Although the parallel-integration model was originally put forward by Calvert and Thesen (2004) merely to explain the variety of cross-modal phenomena, Koelewijn, Bronkhorst, and Theeuwes (2010) argued that it can provide a useful theoretical framework for the investigation of interactions between selective visual attention and multisensory integration. They proposed that supra-threshold stimuli might have a stronger propensity to be integrated automatically, while near-threshold, or weakly effective signals, might require additional attentional resources (cf., Experiment 6).

The conclusion from the literature discussed in the present section is that substrates for multisensory integration exist in various brain areas and at multiple stages of cortical hierarchy. This suggests a complex interplay between multisensory integration and visual object selection, where in some cases spatial selection of visual objects will be biased via a bottom-up mechanism towards salient bimodal (e.g., audiovisual) objects that were created at pre-attentive stages, while in other instances preferential selection of bimodal objects will take place via a top-down mechanism, where multisensory integration occurs at later stages of information processing, once signals in both modalities are separately selected by modality-specific attentional mechanisms.

1.3. Functional interactions between multisensory integration and visual attention

The ability to prioritise the processing of potentially relevant events and objects occurring in the environment is critical for effective cognitive functioning. The past twenty years of research revealed that objects receive preferential perceptual processing not only when they match current or more long-term goals of the observer, but also when they stimulate more than one modality at the same time. As described in Section 1.2, multiple brain areas are recruited by both multisensory integration and selective visual attention, thus it is likely that these two processes influence each other at different stages of information processing. The present section will describe three most intensively investigated interactions between multisensory integration and visual attention: Multisensory enhancement of attentional capture by peripheral visual cues accompanied by non-visual signals (Section 1.3.1), prioritisation of synchronous audiovisual stimuli by visual attention in sequential and visual search tasks (Section 1.3.2), and the influence of top-down attention on multisensory integration (Section 1.3.3). Multisensory enhancement of capture of visuo-spatial attention was the first type of interaction systematically investigated in humans, and showed that the assumed enhanced attentional processing of audiovisual cues over unimodal cues is not always the case. In contrast, investigations into the second and third type of interaction have begun only recently, and more work is required to fully understand the mechanisms underlying these interactions between multisensory integration and selective visual attention.

1.3.1. Audiovisual stimuli as peripheral cues

Animal and human studies demonstrated that when a visual object appears in a spatiotemporal alignment with a signal in another modality, this object will not only trigger an enhanced firing rate of neurons in areas implicated in attention shifting (Macaluso, Frith, & Driver, 2001; Robinson, Bowman, & Kertzman, 1995), but also faster responses (Molholm et al., 2002; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002). These findings motivated investigations into whether multisensory integration can increase the ability of objects to attract exogenous visual attention.

The majority of such employed versions of the spatial cueing paradigm that was originally created by Posner (1980), in which the side where the target appeared is cued in a spatially nonpredictive manner by peripheral unimodal visual or auditory cues, or a

simultaneous presentation of two cues at the same location (bimodal cues). In one such experiment, Santangelo, van der Lubbe, Olivetti Belardinelli, and Postma (2006) instructed participants to discriminate whether a short-duration black triangle pointed up- or downwards. The side of the target presentation was correctly indicated by cues on 50% of all trials. RT data showed that spatial cueing effects were elicited by all cues, both unimodal and bimodal ones. However, in line with previous experiments (e.g., Funes, Lupiáñez, & Milliken, 2007), the cueing effects triggered by unimodal and bimodal cues were comparable in size, indicating that attentional capture in vision is not modulated by multisensory integration. Furthermore, the absence of enhanced attentional capture by bimodal compared to unimodal stimuli was not simply due to a lack of multisensory integration. In an EEG study, Santangelo, van der Lubbe, Olivetti Belardinelli, and Postma (2008) found that amplitudes of the P1 component that audiovisual cues elicited over parieto-occipital sites were larger than the sum of neural responses elicited by unimodal cues (but see Santangelo & Spence, 2008, for an explanation of these results in terms of increased alertness).

While these findings suggested that multisensory integration and exogenous visuo-spatial attention operate independently, Santangelo and Spence (2007a) demonstrated that spatiotemporally aligned bimodal cues can be more effective than unimodal cues in capturing visual attention, but that this occurs only in contexts in which visual attention is strongly focused at another location by a demanding task. In the single-task condition in the discussed study, participants only judged whether a short-duration black rectangle was presented at the top or bottom of the screen (irrespectively of the side of its presentation). In the dual-task condition, participants identified a target-letter embedded in a central RSVP (rapid serial visual presentation) array on 70% of all trials and performed the spatial task on the remaining 30% of all trials. In blocks where participants were performing only the visuo-spatial task, unimodal and bimodal cues again attracted visual attention to a similar extent. In blocks in which two tasks were simultaneously performed, both types of cue led to errors on the RSVP task, implying some degree of attentional capture. However, in the latter, only the audiovisual cues effectively re-oriented attention to the secondary spatial task, as indexed by reliably present RT spatial cueing effects. Notably, the magnitude with which bimodal cues captured visuo-spatial attention was similar across the two conditions, suggesting that attentional processing of the bimodal events was not affected by the varying demands of the task.

Santangelo and Spence (2007a) concluded that the advantage of peripheral bimodal cues over unimodal cues lies in the ability of the former to effectively disengage visuo-

spatial attention from its current focus (cf., Folk, Ester, & Troemel, 2009). Follow-up studies demonstrated that the engagement of visual attention with a demanding task is the critical factor underlying the enhanced attentional processing of bimodal over visual objects: Unimodal cues were shown to be effective in re-orienting attention in contexts in which participants merely passively viewed the central RSVP stream (Santangelo & Spence, 2007b, Experiment 2), but the same cues were ineffective in situations, in which participants needed to monitor a stream of gradually morphing shapes for targets defined by a subtle colour change (Santangelo, Finoia, Raffone, Belardinelli, & Spence, 2008). Interestingly, the ability of bimodal cues to disengage visual attention from a demanding task seems to be critically dependent on whether they convey redundant spatial information. Ho, Santangelo, and Spence (2009) showed that bimodal audiotactile cues re-oriented attention reliably in cases where both components indicated the same side of the visual field, but failed to do so in contexts where the two unimodal stimuli were presented from disparate locations (i.e., tones were presented from speakers located on each side of a screen while vibrations were administered centrally to participant stomach).

The research carried out by Santangelo and colleagues suggested that bimodal objects can attract visual attention more effectively than unimodal objects, but at least in the case of spatiotemporally aligned peripheral objects, this effect might be visible only in situations where the unimodal cues lose their ability to attract visuo-spatial attention. The robustness of attentional capture triggered by such bimodal objects is consistent with the SC-based effect of multisensory integration on attention orienting (see Section 1.1.1.1 for more details; see also Bell, Meredith, Van Opstal, & Munoz, 2005). The SC-mediated mechanism might facilitate the orienting to non-salient peripheral events by extracting the redundancy of spatial information. Importantly, this suggests that bimodal spatiotemporally aligned cues might disengage visuo-spatial attention from a demanding task because they provide spatial information not through one, but two systems of exogenous attention (Eimer & Driver, 2000; McDonald & Ward, 2000, for evidence suggesting at least partial independence of modality-specific systems of exogenous attention). Thus, in attentional capture by this type of bimodal cues, it is difficult, if not impossible (see McDonald, Teder-Sälejärvi, & Ward, 2001), to distinguish the effects of multisensory integration from the effects of cross-modal spatial attention.

Critically, for a peripheral unimodal object to capture attention stronger when accompanied by a spatiotemporally aligned signal in another modality, spatial location of each unimodal stimulus might have to be first selected by the brain (cf., late integration model, Section 1.2.3). Thus, spatiotemporally aligned cues might not have been optimal to

investigate whether it is possible for bimodal objects to capture attention stronger than unimodal objects via a form of multisensory integration which does not require attentional selection to occur. Section 1.3.2 will discuss recent findings which suggested that visual objects that temporally coincide with non-informative signals from other modalities can be automatically integrated at early levels of cortical processing into salient integrated bimodal objects, which in turn enhances their ability to attract shifts of involuntary attention in single-stimulus as well as multi-stimulus contexts (cf., early integration model, Section 1.2.3).

1.3.2. Prioritisation of synchronous audiovisual objects by visual selective attention

The second line of research on the interactions between multisensory integration and visual attention focused on whether the attentional processing of unimodal visual objects is facilitated in contexts where they are coinciding with signals from other modalities. Studies in this are aimed to extend the results of Stein and colleagues (1996), who demonstrated that very basic, low-level perceptual judgements, such as the perceived intensity of a visual stimulus, can be affected merely by concurrent presentation of stimuli to other modalities. The logic that guided research in this area was that the largest effects of multisensory enhancement of visual perceptual representations on visual attention should be visible in the context where the perceptual representation is degraded, in line with the 'inverted effectiveness rule' (see also Section 1.1.1.1). The following sections will describe the existing studies that focused on whether audiovisual synchrony can help overcome the temporal (Section 1.3.2.1) and spatial (Section 1.3.2.2) limitations of visual attention.

1.3.2.1. The role of audiovisual synchrony in sequential selection tasks

The previous sections discussed how common interactions between signals from different modalities are. Thus, it is intuitively correct to assume that in contexts where the perceptual analysis in the primary modality is hampered, e.g., by presence of elevated levels of noise, information in other modalities might be utilised to increase the amount of relevant information (see Section 1.1). The 'freezing effect' (Vroomen & de Gelder, 2000) was one of the first scientific demonstrations showing that attentional selection in vision can be aided by supreme temporal resolution of the auditory system was. Vroomen and de Gelder (2000; Experiment 1) investigated whether the perceptual organisation of an auditory stream can

facilitate performance in a task which places high temporal demands on visual attention. They instructed participants to detect within a series of successive, rapidly changing trials which set of four dots presented in visual displays align to form a cross. Crucially, each display was accompanied by a task-irrelevant tone. In a condition where the target was synchronised with a deviant tone embedded in an array of identical tones, the accuracy of target detection was higher, and a subjective experience of prolonged exposure of the visual target was also reported (i.e., 'freezing' of the target display). In follow-up experiments, Vroomen and de Gelder (2000) showed that the observed performance facilitation was driven by the deviant tones increasing the ability of participants to segregate the visual target from the stream it was embedded in, rather than the sound serving as a temporal cue indicating the target onset (Experiments 2 through 4). This account was further supported by a lack of improvement found in a context where the tone was less abrupt or part of a melody (Experiments 5 and 6).

These findings could be regarded as evidence that auditory signals can aid visual targets to segregate from a rapid stream through enhancement of their perceptual representation, but it is unclear to what extent similar results could be observed when the visual target would be synchronised with an absence of a tone within the auditory stream. Using the 'attentional blink' (AB; Raymond, Shapiro, & Arnell, 1992; Shapiro, Arnell, & Raymond, 1997) paradigm, Olivers and van der Burg (2008) provided strong support for the notion that pairing a non-informative tone with a visual stimulus can increase the propensity of the latter to be attentionally selected by strengthening its neural representation. Attentional 'blink' describes a phenomenon in which identification of a second target (T2) in a RSVP array is usually impaired if this target is presented within approximately half a second of the first target (T1). While numerous theories have been put forward to explain the AB phenomenon, the consensus is that the nature of this effect is post-perceptual, i.e., the effect is driven by an impairment of memory and/or response selection processes. This account is substantiated by results showing that more salient T2 targets, such as physically dissimilar stimuli or the participant's own name, are resistant to the described impairment (Chun & Potter, 1995; Di Lollo, Kawahara, Shahab Ghorashi, & Enns, 2005; Olivers, 2007; Shapiro & Raymond, 1994). Crucially, Olivers and van der Burg (2008) showed that pairing T2 with concurrent task-irrelevant auditory distractors can have the same effect (Experiment 1). While follow-up experiments revealed that this tone-induced improvement of performance cannot be explained by alerting (Experiment 2), it seems to be strongly dependent on cross-modal endogenous temporal attention (Experiment 3): In contexts where the sound coincided with distractors on most (80%) trials, participants were shown to be

actively suppressing the tones, while no evidence of this was found in Experiment 1, where auditory events were predictive of the T2 onset. In spite of this top-down inhibition, attenuated but still reliably present performance improvements were observed in such conditions on trials, in which the sound was presented concurrently with the visual target. These findings provided the first strong evidence that audiovisual synchrony can create a bottom-up bias in visual object selection (cf., Desimone & Duncan, 1995) by enhancing the neural representation of a primary stimulus. However, what is important, paradigms involving serial stimulus presentation cannot provide insights into whether multisensory integration also creates a bottom-up bias towards visual objects paired with non-visual signals in multi-stimulus contexts.

1.3.2.2. The role of audiovisual synchrony in spatial selection tasks

If pairing a noninformative sound with a visual object can strengthen the neural representation of this object, the increased salience of such a stimulus should in principle also enable it to be preferentially selected in contexts where multiple simultaneous objects compete for attentional selection. Using two different paradigms, i.e., a visual search task and a temporal order judgement task, van der Burg, Olivers, Bronkhorst, and Theeuwes (2008a, 2008b) provided initial evidence that multisensory enhancement of visual salience can affect selection of objects in multi-stimulus contexts.

In the first study, van der Burg et al. (2008a) demonstrated that slow and inefficient visual search carried out in a cluttered dynamic display is dramatically improved when the target is paired with a task-irrelevant, spatially nonspecific tone. Critically, while the audiovisual target seemed to subjectively pop out of such arrays (‘pip-and-pop’ effect; see also Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2009, for similar effects found for visuo-tactile pairings), the observed search benefit was shown to be strongly dependent on the degree of co-occurrence between sound and targets vs. sound and distractors (Experiment 4), suggesting that the sound in this paradigm was predominantly utilised as a temporal cue (see Chapter 2 for more details). Some evidence that tone-induced enhancements of visual salience might have also contributed to the observed search facilitation was provided by Experiment 5, in which visual targets were always presented without tones: Overall, larger search costs were now observed on those trials in which targets were preceded by distractors accompanied by tones compared to purely visual distractors. However, in blocks where the time interval between the distractor and target was

short (i.e., 200 ms), search was more strongly disrupted by visual relative to audiovisual distractors, thus precluding drawing strong conclusions from this study in respect to the role of audiovisual synchrony as a source of salience-based bias in visual object selection.

To provide a more direct test of whether multisensory integration can enhance visual attention capture via a bottom-up mechanism, van der Burg et al. (2008b, Experiments 1 and 2) used a temporal order judgement (TOJ) task. In one of the experiments (Experiment 1), participants judged which of two lateral target dots appeared first. Each dot was accompanied by a task-irrelevant distractor dot which could change colour prior to appearance of the first target dot. As temporal order judgements are known to be biased by attentional orienting (Shore, Spence, & Klein, 2001), they are suitable to assess if audiovisual salience can enhance the ability of visual objects to capture attention in multi-stimulus contexts. Results showed that temporal judgements were biased towards the location of the distractor dot, but only in blocks in which it changed colour together with a presentation of a spatially diffuse tone. While these findings are in line with an enhancement of spatial selection of visual objects by audiovisual salience, the influence of response bias on TOJ could not be ruled out in this experiment (see Chapter 2 for more details). When a less biased measure of temporal perception was used (simultaneity judgements; van der Burg et al., 2008b, Experiment 3), audiovisual distractors again affected the temporal judgements, but the influence of audiovisual integration on visual attentional capture could not be assessed due to a lack of an unimodal distractor condition against which this effect could be compared. Thus, while the studies of van der Burg and colleagues (2008a, 2008b) have suggested that in multi-stimulus contexts the ability of visual objects to capture attention when paired with non-visual signals might be enhanced via a salience-based mechanism, they failed to do so in a convincing manner.

1.3.3. Influence of top-down attention on multisensory integration

Multisensory integration can occur at various stages of cortical processing, which makes the integrative process sensitive to various modulations by selective attention. Evidence from behavioural, electrophysiological and neuroimaging studies (Calvert et al., 2000; Giard & Peronnet, 1999; Vatakis & Spence, 2007) has recently highlighted several cognitive factors that can influence the presence and the strength of combining stimuli across modalities.

One of the most important cognitive factors influencing multisensory integration is the ‘unity assumption’ which defines the degree to which an individual, consciously or

unconsciously, assumes that two signals originate from the same object or event. If two stimuli are regarded as emerging from a single source, enhanced effects of multisensory integration can be observed. For example, Vatakis and Spence (2007) demonstrated that temporal discrimination of the onset of two speech stimuli is more difficult when the two stimuli are matched (e.g., male face presented together with male voice) than when they are mismatched (e.g., male face paired with female voice). In this context, endogenous attention might be even more important for multisensory integration, as it modulates the interaction of stimuli from different modalities by altering perceptual representations and associated neural responses. In a recent fMRI study, Fairhall and Macaluso (2009) showed that directing spatial attention to matching, as opposed to nonmatching, lip movements while listening to spoken sentences results in enhanced activation in multiple brain areas, including subcortical (i.e., SC) and cortical (i.e., V1 and STS) structures. Additionally, susceptibility to the so-called 'McGurk illusion', where a lip movement paired with an incongruent auditory phoneme results in an illusory auditory percept, can be reduced in cases where endogenous attention is focused on a perceptually demanding visual or auditory task (e.g., Alsius, Navarra, Campbell, & Soto-Faraco, 2005). These findings provided evidence that multisensory integration and visuo-spatial attention can jointly affect perception of a fused audiovisual stimulus.

However, it needs to be noted that speech belongs to the class of 'social stimuli' (Lieberman, 1996) and as such might recruit unique neuro-cognitive mechanisms, which limits the generalisability of the conclusions that can be drawn from the studies discussed here about the degree of modulation of multisensory integration by cognitive factors. Thus, it is important that similar effects were found for combinations of more basic stimuli (e.g., beeps and flashes). Talsma, Doty, and Woldorff (2007) employed the event-related potentials (ERP; see Section 1.4) technique to investigate whether early-latency effects of multisensory integration (for details, see Section 1.4.2) are modulated by endogenous attention. In their study, an RSVP stream of letters presented centrally above the point which participants fixated was accompanied on some trials by visual (horizontal gratings presented below fixation), and auditory (spatially diffused high-frequency tones) stimuli presented separately or together. In separate blocks, participants were instructed to (i) attend only to the RSVP stream to detect infrequent target-digits, (ii) attend to all the visual, auditory and audiovisual objects to detect rare dips in intensity occurring halfway through the stimulus duration, (iii) attend to visual objects and visual components of bimodal objects, or (iv) attend to auditory objects and respective components of bimodal objects. Talsma et al. (2007) found early-latency 'superadditive' neural responses (see also Fort et al., 2002a;

Giard & Peronnet, 1999) to redundant bimodal targets and the associated facilitation of target detection, but only in blocks where visual and the auditory modality were simultaneously attended. In blocks where attention was directed to the RSVP stream and thus away from the bimodal objects, the neural responses they elicited were 'subadditive' (i.e., smaller than the sum of unimodal responses) and target detection was no longer facilitated.

It is to be noted that in blocks, where subjects attended only to the visual components of bimodal objects, the effects of audiovisual redundancy were visible, only at longer latencies, i.e., an enhanced frontal negativity was found at around 250 ms post-stimulus. Interestingly, similar enhancements, notably, in the range of the 'late auditory processing' ERP component, were found to characterise another attention-based multisensory phenomenon, i.e., the 'cross-modal spread of attention' (Busse, Roberts, Criss, Weissman, & Woldorff, 2005, but see also Fiebelkorn, Foxe, & Molholm, 2010). Busse and colleagues (2005) demonstrated that in contexts where a central task-irrelevant sound is paired with an attended lateral checkerboard, this sound triggers an enhanced late processing negativity, indicative of attentional processing of the auditory distractor, but no such enhancement is observed in contexts where the visual stimulus is unattended. This distinctive pattern of neural activity, triggered by a cascade of top-down (attention to the visual event) and bottom-up (temporal proximity of visual and non-visual signals) influences, seems to reflect a separate class of cross-modal interactions: The initial attention-based increase of neural activity triggered by the visual target spreads automatically to non-visual cortices, where it results in an attention-like neural response to a concurrent event, in spite of the fact that the latter is spatially misaligned with the primary stimulus and irrelevant to the task-at-hand.

While the evidence discussed above clearly suggests that numerous top-down factors mediate or strongly modulate multisensory integration, a fundamental question in this area remains unanswered: Can bimodal objects be selected preferentially over unimodal objects due a stronger match with a top-down multi-feature attentional template (cf., Duncan & Humphreys, 1989)? Semantic congruence might be of particular importance in cases where multisensory integration might serve as mechanism of top-down control of over selection of objects in multi-stimulus contexts. Pairing a visual objects with a semantically congruent non-visual stimulus is known to facilitate object recognition (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Molholm, Ritter, & Javitt, 2004; Smith, Grabowecky, & Suzuki, 2007). More importantly, Iordanescu, Grabowecky, Franconeri, Theeuwes, and Suzuki (2010) showed that semantic congruency can also improve visual search. Saccade

latencies to target objects were shorter in conditions where a target-present array (e.g. an array with a dog picture) was accompanied by a target-congruent sound (e.g., a barking sound) relative to an incongruent-sound (e.g., meowing sound) or a no-sound condition. These results suggest that multisensory integration might potentially be a source of top-down bias in selection of objects in multi-object contexts. However, the study of Iordanesco et al. (2010) does not allow any direct insights into the nature of the mechanisms that control search for bimodal objects, as their results could be explained equally well by guidance of attention by an integrated audiovisual object template and by separate modality-specific features guiding visual selection. Establishing whether search for targets defined by feature conjunctions is controlled in the same effective task-dependent fashion irrespective of where these targets are defined within a single modality (cf., Kiss et al., 2013) and in contexts where targets are defined across modalities is also important for the validity of the visual attention theories that argue for the dominant role of the top-down mechanisms (Bundesen, Habekost, & Kyllingsbaek, 2005; Desimone & Duncan, 1995; Wolfe, 2007) in the control of object selection in multi-stimulus contexts.

1.3.4. Initial theoretical framework

Within the past couple of decades converging behavioural and neuroimaging evidence was provided in support of the idea that multisensory integration and selective attention can influence each other. As described in Section 1.3.2, there are instances, in which temporally coincident signals from different modalities are integrated in a pre-attentive and effortless manner into a salient emergent multimodal object which has an increased ability to attract exogenous visual attention (Olivers et al., 2008). In contrast, the literature discussed in Section 1.3.3 strongly suggests that certain forms of cross-modal interactions are dependent on endogenous attention. How can these disparate results concerning the direction of interaction between selective attention and multisensory integration be reconciled?

Recently, Talsma, Senkowski, Soto-Faraco, and Woldorff (2010) proposed an initial theoretical framework to describe and explain the factors that play a crucial role in this interplay. According to their model, the ‘complexity’ of the multisensory environment determines whether multisensory integration will require attentional resources or occur automatically. Talsma et al. (2010) defined the ‘environment complexity’ as the degree of ongoing competition occurring between stimuli within each modality. In particular, the probability of effortless, automatic multisensory integration is higher in task contexts in which the competition between stimuli in the other modality is low, e.g., in contexts where

the events are rare. Infrequency of these stimuli should increase their bottom-up salience and trigger a neural response strong enough to be automatically associated with a response to a concurrently presented object or event in the primary modality. For example, this account is supported by the results of Olivers et al. (2008), which demonstrated that sparse auditory stimulation can increase the ability of a concurrent visual object to be selected from a rapidly changing array.

In their framework, Talsma and others (2010) contrasted this context with one, in which there were multiple stimuli present in close succession in each modality. The competition for processing resources that occurs in such circumstances decreases the bottom-up salience of stimuli in the other modality, which necessitates the presence of endogenous attention for integration of appropriate, task-relevant signals and effective processing of the resulting integrated multimodal object. Support for this account is provided by the study of Talsma et al. (2007), in which the suppression of neural response to unattended co-occurring visual and auditory stimuli could be explained by depletion of processing resources due to focusing attention on the concurrent RSVP task. In other words, in contexts where competition decreases the perceptual salience of the stimuli presented in the other modality, unimodal signals might have to be first separately selected by respective within-modal attentional mechanisms in order for them be integrated into a multimodal object. Notably, the notion, according to which attention plays a role of 'glue' that integrates appropriate features in multi-stimulus environments, is one of the core assumptions of FIT (see Section 1.1.2).

The framework proposed by Talsma and colleagues was the first attempt to integrate the plethora of findings that represent two very different facets of the interplay between multisensory integration and selective attention. While it would be highly beneficial for the proposed hypotheses to be tested through systematic manipulation of the degree of competition between stimuli within the other modality, their plausibility seems to be supported by their resemblance to the tenets of the 'perceptual load theory' of visual attention (Lavie, Hirst, Fockert, & Viding, 2004; Lavie, 2005, 2010; Lavie & Tsal, 1994). According to this model, if the competition for processing resources among visual stimuli is low, task-irrelevant stimuli will be automatically processed. However, if processing resources are depleted by a perceptually demanding task, the task-irrelevant distractors will be filtered out at early stages of information processing. Extrapolating to multisensory contexts, it can be assumed that when the task-at-hand does not impose high demands for perceptual resources, these resources will be automatically diverted to the processing of the co-occurrence of signals from different modalities, producing a salient multimodal objects

that is preferentially selected when presented among purely visual objects. The idea of Talsma and colleagues, according to which the stimulus delivery rate (i.e., the length of the interval between the onset of the present and the successive event) is critical in determining whether the integration of signals across modalities takes place in a pre-attentive manner (see Fujisaki, Koene, Arnold, Johnston, & Nishida, 2006, for results showing that stimuli presented faster than one every 250 ms fail to do so) is consistent with the perceptual load theory.

Overall, Talsma and colleagues (2010) were the first to aim to identify a single factor that might be of critical importance for the directionality of the influence between selective attention and multisensory integration. The most important contribution of this model to the research presented in this thesis is highlighting the need for investigation of the interplay between visual attention and cross-modal integrative processes in ecologically valid contexts where multiple stimuli compete with each other for selection.

1.4. Methodological approach

A method well suited for investigations of the influences of multisensory integration on visual attention is one involving a combination of performance and event-related potentials measures. The ERP technique involves recording the brain activity at the scalp (i.e., electroencephalogram, EEG) and dissociating the signal associated with processing of specific events from the background noise. Together with behavioural indices of attentional selection, ERPs can provide an important insight into whether audiovisually-induced enhancements of spatial selection of visual objects accompanied by irrelevant tones are already visible at stages concerned with selective perceptual processing. In the context of search for objects defined as conjunctions of visual and auditory features, ERPs, as a direct measure of neural processing of items that do not require an overt response, can reveal whether and, if so, via which mechanisms (i.e., onset latency effects or just amplitude effects) attentional selection of task-irrelevant cues changes depending on how many features they share with the target. Importantly, in this area, ERPs can reveal the stage at which selective processing is already controlled by integrated audiovisual object templates. The aim of the following sections is to describe the biophysical basis of the ERP technique and to provide an overview of ERP components and ERP effects associated with multisensory integration and visual attention.

1.4.1. The neural basis of event-related brain potentials

The EEG signal is generated by synchronised activation of large populations of pyramidal cortical neurons. There are two main types of electrical activity exhibited by neurons, i.e., action potentials and postsynaptic potentials. The voltages recorded by scalp electrodes are believed to be generated by postsynaptic potentials of pyramidal cells (Luck, 2005). Action potentials, i.e., discrete voltage spikes that travel along the axon of a neuron, are not usually detectable at the scalp level. This is due to a very short duration (around 1 ms) of these spikes, as well as the physical arrangement of axons, which leads, in cases where they are not in full synchrony, to the cancellation of spikes produced by one neuron by spikes from another. In turn, postsynaptic potentials are generated when an excitatory neurotransmitter is released from the axons of the presynaptic cell and binds to the receptors of the postsynaptic neuron, altering its membrane potential: Positive ions enter the cell at the level of apical dendrites and create a net negativity outside the neuron in the region of the dendrite. The electrical circuit is completed by the current flowing out of the body cell at the basal dendrites, resulting in a net positivity in this region.

Positive and negative discharges of current, which are separated in space, create a 'dipole'. A dipole that is a product of a single neuron is too miniscule to be recorded by a scalp electrode, but this becomes possible if dipoles created by larger ensembles of neurons summate. Two conditions need to be fulfilled in order for the summated voltages to be recordable at the scalp level. First, the same postsynaptic activity must be generated at approximately the same time across thousands of neurons. Second, the dipoles from single neurons need to be spatially aligned; if the orientation between different dipoles exceeds 90 degrees they will cancel each other out. Such spatial alignment is most typical among pyramidal neurons which orientation is perpendicular to the cortex surface. Importantly, the summation of dipoles is further complicated by the fact that the cortex is not flat, but folded. However, the alignment of multiple smaller dipoles typically approximates a single dipole, as the orientations of these dipoles are averaged ('equivalent current dipole'; Luck, 2005).

Neural processes which are time-locked to specific events can be extracted from the overall EEG activity through a simple averaging process. The activity that is associated with a stimulus will be triggered by each presentation of the stimulus and will contribute to the average, while the activity that is connected to processes not related to the stimulus will be gradually attenuated with the increasing number of trials being averaged. This process of averaging across hundreds of stimulus repetitions leads to creation of EEG time series, which represents a typical neuro-cognitive response to an event of interest and which is

composed of a sequence of positive and negative voltage deflections, also known as ERP components. An ERP component is usually labelled with a letter describing its polarity ('N' for negative, 'P' for positive) and a number that indicates its ordinal position (e.g., 'N1' means the first negative peak) or latency (e.g., 'P300' means a positive component that is usually observable around 300 ms after onset of the stimulus of interest). The remaining parts of this section will describe the components and effects associated with multisensory integration and visual attention.

1.4.2. ERPs related to multisensory integration

Due to its high temporal resolution, EEG was one of the first methods that provided evidence that some instances of cross-modal interactions can take place through feedforward convergence of signals from different modalities. Research in this area was motivated by questions whether the increased speed of responses to redundantly defined (i.e., requiring the same response) bimodal stimuli that was reported by the behavioural studies published at the time (Bertelson & Radeau, 1981; Miller, 1982) was accompanied by neural modulations at early (i.e., perceptual) or later (i.e., response-related) stages of information processing.

In a pioneering study in this area (Giard & Peronnet, 1999), participants had to discriminate which of two different objects was presented on each trial by pressing one of two keys. Each of two target objects was defined by a visual feature (e.g., deformation of a circle to a horizontal ellipse; V) and an auditory feature (e.g., tone of 540 Hz; A). On trials in which both features were presented together (e.g., circle deformation accompanied by a 540-Hz tone; AV), responses were faster and more accurate compared to stimuli with task-relevant features presented alone. The critical measurement in this study involved a comparison of the neural activity triggered by the redundantly defined audiovisual objects with the responses triggered by visual and auditory target-defining signals presented in isolation: If, at a given information processing stage, features from different modalities are processed *independently* despite their simultaneous presentation, this should be reflected by a comparable level of neural activation when two features are presented concurrently and separately:

$$\text{Response to AV} = \text{Response to V} + \text{Response to A}$$

In contrast, any activity that exceeds the sum of unimodal responses would indicate integration of the two signals:

Response to AV > Response to V + Response to A

Giard and Peronnet (1999) found ERP differences between the audiovisual stimuli and the sum of unimodal stimuli over occipital sites as early as 40–45 ms post-stimulus. Notably, the redundantly defined bimodal objects were shown to trigger a whole range of dissociable ERP effects, from early enhancements, with likely sources localised in the visual and auditory sensory cortices, to late, modality-nonspecific effects around 180 ms post-stimulus, with likely generators in fronto-temporal cortex.

The validity of this 'additive model' in uncovering various overlapping subcomponents in a brain response to bimodal events stems from the 'law of superimposition of electrical fields', which states that potentials from separate current sources in a conductor sum linearly. Hence, the interaction term that results from subtraction of two unimodal ERPs from the bimodal ERPs will only reflect only the volume conduction effects triggered by cross-modal interactions, as conduction effects related to unimodal responses will be eliminated. The 'additive' model was first implemented to study ERPs in cats by Berman in the early 1960's, and more recently was employed by Barth, Goldberg, Brett, and Di (1995) to identify brain areas selectively responsive to audiovisual stimulation. Since then it has been extensively used to study cross-modal interactions in the human brain (e.g., Foxe, Morocz, Murray, Higgins, Javitt, & Schroeder, 2000; Murray, Molholm, Michel, Heslenfeld, Ritter, Javitt, & Schroeder, 2005; Talsma & Woldorff, 2005).

However, the early-latency ERP interactions reported by Giard and Peronnet (1999) were criticised by Teder-Sälejärvi, McDonald, Di Russo, and Hillyard (2002), who showed that these effects might reflect an artefact resulting from processes active in all three conditions, e.g., anticipatory effects. The activity that is related to stimulus expectation is usually visible as the contingent negative variation (CNV) component, i.e., a slow potential arising before each stimulus and continuing after its presentation. Teder-Sälejärvi and colleagues (2002) demonstrated that when control measures are employed (i.e., a high-pass filter with a 2Hz cut-off frequency and a change of the reference baseline period) to prevent such spurious ERP activity, the earliest neural activity arising due to cross-modal interactions is visible only around 120-130 ms post-stimulus. However, Fort et al. (2002a) subsequently shown that ERP responses to bimodal stimuli are strongly dependent on the requirements imposed by the task, with anticipatory effects being characteristic of

paradigms in which stimuli are presented rapidly and which place strong emphasis on the speed of the responses. Fort and colleagues (2002a) provided evidence that in cases where stimuli are presented for 250 ms and are accompanied by relatively long and variable interstimulus intervals (between 1400 and 3300 ms), ERP enhancements are again observed over occipital sites as early as 50 ms post-stimulus (cf., Giard & Peronnet, 1999; see Murray et al., 2005, for similar effects over central sites after audiotactile stimulation).

Importantly, the early-latency neural interactions between visual and audiory stimuli revealed by Giard and Peronnet (1999) were shown to be sensitive to task requirements (Fort et al., 2002a, 2002b). In one study, participants identified three objects defined by auditory or visual features alone or by a combination of nonredundant auditory and visual features (Fort et al., 2002b). In another study, participants were instructed to detect any of the three objects, irrespective of whether presented in a unimodal or bimodal form (Fort et al., 2002a). Both studies used simple shape- and pitch-based visual and auditory stimuli similar to the ones employed by Giard and Peronnet (1999). The shortening of response latencies and early-latency sensory modulations were visible in the detection task, but not in the non-redundant categorisation task, which can be explained by the requirement posed by the latter task for processing of both features in a non-redundant object discrimination task. Notably, in all three tasks (Fort et al., 2002a, 2002b; Giard & Peronnet, 1999), similarly enhanced neural responses to bimodal versus unimodal stimuli were found over the right fronto-temporal sites around 150 ms post-stimulus. Due to the robustness of this ERP effect across different tasks, Besle, Fort, and Giard (2004) attributed it to a process reflecting a 'more general integration function', such as detection of audiovisual synchrony, and, in line with the existing literature (Bushara, Grafman, & Hallett, 2001), localised its possible neural generator in the right insula.

To conclude, except for isolated cases (e.g., Santangelo et al., 2008), the ERP technique was used in the multisensory research predominantly to investigate the influence of task demands on neural responses to cross-modal interactions across different stages of the information processing. The ERP studies conducted by Giard and colleagues (Giard & Peronnet, 1999; Fort et al., 2002a, 2002b) provided the first evidence that signals from different modalities can interact at very early stages of cortical processing in the human brain. The timing of these ERP effects is inconsistent with the potential involvement of feedback projections (Foxe & Schroeder, 2005) from heteromodal brain areas, thus suggesting they are supported by connections between early sensory cortices that operate in a direct lateral fashion (Falchier et al., 2002) or in a feed-forward fashion that is mediated by the pulvinar nucleus of thalamus (see Cappe et al., 2009, for a review). However, the fact

that these early-latency cross-modal interactions were shown to be strongly dependent on the task-at-hand did leave open the questions whether multisensory integration can bind stimuli from different modalities into a single salient bimodal object via an automatic mechanism at early stages of cortical processing, and whether it can provide this integrated object with a competitive advantage at later stages of cortical processing at which multiple events compete with each other for attentional selection.

1.4.3. ERPs related to visual attention

Over the past 40 years, ERP components related to visual attention were intensively investigated. Due to their exceptional temporary resolution, ERPs were invaluable in elucidating the various mechanisms through which visual attention achieves its selective nature, as well as in describing the stages of information processing at which it operates. Section 1.4.3.1 focuses on the early sensory components that are affected by visuo-spatial attention. Sections 1.4.3.2 and 1.4.3.3, respectively, discuss components that represent mechanisms involved in attentional orienting and attentional selection of candidate target items among distractors.

1.4.3.1. Sensory 'gating' mechanisms revealed by ERPs

The P1 and N1 components, regarded as reflecting early stages of sensory processing, are the first ERP components that are affected by attentional selection. They are predominantly influenced by physical properties of the stimuli, but visuo-spatial attention modulates them in a comparable way. The P1 component is a positive deflection that typically peaks around 80-130 ms after stimulus onset, which amplitudes are the largest over lateral posterior sites. The likely neural generators of this component are the middle occipital gyrus and the fusiform gyrus (Russo, Sereno, Pitzalis, & Hillyard, 2001) with possible contributions from several other visual areas that are active during early stages of information processing. Due to its origins in the extrastriate cortex, P1 is strongly affected by physical stimulus properties, such as brightness or contrast. However, it is not sensitive to most cognitive processes, except for arousal (Vogel & Luck, 2000) and direction of spatial attention (for a review, see Hillyard, Vogel, & Luck, 1998). The N1 component is the first negative component elicited during visual cortical processing and it is composed of several subcomponents. The earliest N1 peaks around 100-150 ms after stimulus onset and is typically observed over anterior electrode sites. Two posterior N1 subcomponents peak

around 150-200 post-stimulus, where one is visible over parietal and the other over lateral occipital sites. The lateral occipital N1 subcomponent is thought to reflect discriminative processes, as it is typically larger in tasks involving stimulus discrimination rather than stimulus detection (Hopf, Boelmans, Schoenfeld, Heinze, & Luck, 2002; Luck, Woodman, & Vogel, 2000).

Numerous studies have shown that P1 and N1 are influenced by the direction of spatial attention (for a review, see Mangun, 1995), both sustained (i.e., participants focus their attention continuously on one specific visual region in space) and transient (i.e., the likely location of the target is indicated on a trial-by-trial basis by a symbolic cue). Importantly, the amplitude of these components was shown to be enhanced for stimuli presented at attended, as opposed to unattended locations, providing evidence that early perceptual processing is facilitated by the direction of spatial attention (Eimer, 1994a; Hillyard & Anllo-Vento, 1998; Luck & Hillyard, 1994; Mangun & Hillyard, 1991). The majority of such studies used spatially informative symbolic cues and explained these amplitude enhancements in terms of benefits of preparatory attentional signals (i.e., endogenous attention). However, the P1 component was also shown to be sensitive to non-predictive exogenous cues, with the pattern of ERP modulations closely resembling the behavioural effects following such cues (Eimer, 1994b; McDonald, Ward, & Kiehl, 1999). Short cue-target intervals elicit P1 amplitude enhancement (and attentional benefits), while long cue-target intervals result in amplitude attenuation (and 'inhibition of return'; see Section 1.2.1). These findings provide corroborating evidence that the two discussed ERP components in fact reflect two separate mechanisms: While P1 reflects a mechanism involved in the control of enhancement of sensory processing, N1 is linked to active engagement of attention at specific location (Mangun & Hillyard, 1991; Mangun, 1995; but see Luck & Beach, 1998, for a different interpretation).

The findings of spatial modulations of sensory ERPs are often regarded as evidence in support of the early selection account of selective attention, where selective processing grants access to limited-capacity serial processing to visual stimuli during early stages of information analysis, i.e., on the basis of their physical features, such as location in space. However, these attentional modulations might merely reflect the fact that the neural response that an object elicits in modality-specific cortices (i.e., similar attentional modulations were also found for early sensory auditory components; Hillyard, Hink, Schwent, & Picton, 1973) can be selectively enhanced when attention is directed to location of this object. Research discussed in the sections below demonstrates that the growing consensus is that attentional selectivity can operate at different stages of information

processing, with ERP effects of selective attention visible between 200 and 800 ms post-stimulus (e.g., Eimer, Van Velzen, & Driver, 2002; Eimer, 1996).

In fact, research on attentional modulations of sensory ERPs provided evidence against one of the tenets of the early selection model, i.e., a strictly sequential order of stages of sensory information processing. Behavioural studies conducted by Spence and Driver (1996, 1997) revealed presence of spatial links between vision, audition and touch in both endogenous and exogenous attention systems: Orienting attention to a location in one modality was shown to enhance processing of events in this location in other modalities. The ERP technique was particularly useful here as it revealed that this form of cross-modal attentional facilitation is based on links between vision, audition and touch that are hardwired in the brain, as opposed to being driven by more strategic, top-down factors (Eimer & Driver, 2000; Eimer & Schröger, 1998; Eimer, 1999; Kennett, Eimer, Spence, & Driver, 2001; Teder-Sälejärvi, Münte, Sperlich, & Hillyard, 1999)

In one of these studies, Eimer and Driver (2000) instructed participants to direct their attention to one side of the visual field, in which they were to detect infrequent targets within touch, and ignore targets presented at the unattended side and all stimuli presented in vision. The visual stimuli were clusters of LED lights at two opposite sides of the visual field, while tactile stimuli were presented via stimulators attached to the respective index fingers, in close proximity to the LED lights. Results showed that when subjects were instructed to attend to the tactile modality, task-irrelevant visual stimuli presented close to the location of attended tactile targets elicited P1 and N1 components that were larger than when the same stimuli were presented to the unattended side, indicative of cross-modal spatial synergies modulating perceptual stages of information processing. Similar cross-modal links were found between vision, audition and touch in exogenous attention, with peripheral, spatially uninformative events in one modality enhancing early ERP components in response to subsequent stimuli in the primary modality presented at the same location (Kennett et al., 2001; McDonald & Ward, 2000).

From the point of view of the research presented in this thesis, the importance of studies investigating attentional modulations of the P1 and N1 components lies in demonstrating the existence of cross-modal spatial synergies through which selective attention leads to enhanced neural perceptual processing at early stages of cortical processing in different modalities. Notably, in all cross-modal studies in this area (e.g., Eimer & Driver, 2000, 2001; Eimer & Schröger, 1998), enhancements of the early ERP components were larger for stimuli in the relevant compared to the irrelevant modality, arguing against an account where spatial selection that arises in the primary modality

spreads via direct 'horizontal' connections across otherwise separate systems (Spence & Driver, 1997). The next section discusses research suggesting that these perceptual modulations are the result of coordination of spatial attention across modalities (Eimer et al., 2002; see also Farah, Wong, Monheit, & Morrowt, 1989).

1.4.3.2. Attentional preparatory states revealed by ERPs

Evidence that the facilitatory effects of cross-modal attention on perceptual processing are the *result* of the supramodal nature of attentional control mechanisms was provided by studies which investigated the mechanisms involved in the initiation and maintenance of spatial orienting. Although this area is dominated by neuroimaging studies (see Section 1.2.2), the ERP technique affords valuable insights into the temporal order of these processes (Eimer et al., 2002; Harter, Miller, Price, LaLonde, & Keyes, 1989; Nobre, Sebestyen, & Miniussi, 2000).

In one of such studies, Eimer and colleagues (2002) instructed participants to direct attention to the left or right side to detect rare auditory or tactile targets. To reveal the mechanisms involved in orienting of spatial attention, the neural activity triggered by a symbolic cue that directed attention to one side of the visual field was recorded. The 'left-' and 'right-pointing' cues were identical in shape (i.e., triangles pointing in opposite directions) and the target side was indicated by the pointing direction of a triangle of a particular colour (e.g., red), thus preventing contamination of ERP correlates of mechanisms involved in attention shifting from contamination by sensory differences. Similarly to purely visual experiments (e.g., Nobre et al., 2000), the analysis of ERP waveforms elicited by these left- and right-pointing cues in both auditory and tactile tasks revealed a sequence of two lateralised components, i.e., ADAN ('Anterior Directing Attention Negativity'), a frontal negativity with a latency of 300-400 ms, was followed by LDAP ('Late Direction Attention Positivity'), a posterior contralateral positivity with a latency of 500 ms. The ADAN is assumed to reflect activation of the dorsolateral frontal control mechanisms involved in programming and initiating attentional shifts (Nobre et al., 2000; Posner & Petersen, 1990), while the later LDAP is a correlate of preparatory changes in excitability of the ventral occipitotemporal cortex involved in processing of a stimulus expected at a specific location (Harter et al., 1989). In some studies (e.g., Harter et al., 1989; Yamaguchi, Tsuchiya, & Kobayashi, 1998) another component was found to precede ADAN and LDAP: The EDAN ('Early Directing Attention Negativity') is a negative deflection over posterior sites visible around 200 ms post-stimulus contralaterally to the cued side that is associated

with directing of the attention shifts (but see Van Velzen & Eimer, 2003, for evidence suggesting that this component is a lateralised attentional response to laterally presented visual precues, and is thus not directly involved in top-down attentional control).

Critically, Eimer and colleagues (2002) showed that the ADAN and LDAP components have strikingly similar properties (i.e., in respect to latency and scalp distribution) during attention shifts towards the expected location of upcoming auditory or tactile events. While these results suggest that attention shifts are controlled by a supramodal mechanism, other results found in this study were not entirely in line with this account: Attentional modulations of sensory components were the largest for the primary modality, and irrelevant tactile events showed no early ERP enhancements (i.e., touch was 'decoupled' from multimodal spatial attention; see also Eimer & Driver, 2000). Eimer et al. (2002) proposed that, while spatial selection is controlled by a supramodal mechanism, its effects on perceptual processing interact with task-dependent tonic activation within each modality. Additionally, the mechanisms reflected by the ADAN and LDAP also seem to be based on distinctive spatial frames of reference: The programming and initiation of attention shifts (i.e., ADAN) is implemented via a vision-independent egocentric and somatotopic frame of coordinates, while preparatory changes in visual cortical excitability (i.e., LDAP) are contingent on a visually defined frame of reference (Van Velzen, Eardley, Forster, & Eimer, 2006).

The ERP evidence discussed here provides evidence that the mechanisms that control shifts of attention to spatial locations are similar irrespective of the relevance of the modality to the task-at-hand: Supramodal control signals, likely generated in higher-level brain areas (e.g., frontal cortex; Graziano, Yap, & Gross, 1994), appear to modulate early perceptual analysis in low-level sensory regions. The importance of these findings lies in providing a novel perspective on spatial selection, where locations in space are represented in a largely supramodal fashion.

1.4.3.3. Selection of targets among distractors: The N2pc component

The majority of ERP studies described in the previous section employed paradigms in which the effects associated with attentional shifts were revealed by recording brain potentials to symbolic cues indicating the likely location of a single target. However, in everyday life people often do not have the advance information about the location of the currently sought item and thus target objects need to compete for the attentional selection with other stimuli

which are simultaneously present in the visual field. In such circumstances, attention serves to resolve competition on behalf of the stimuli matching current behavioural goals while filtering out all others. The visual search paradigm (see Section 1.1.2, for more details), in which subjects search for pre-defined targets among multiple distractors, is invaluable in investigations of the role of selective attention as a mechanism resolving the competition in multi-stimulus contexts. In this paradigm, the N2pc component, an enhanced negativity that arises approximately 200 ms after onset of the search display over posterior electrodes contralateral to the side of target presentation (Eimer, 1996; Luck & Hillyard, 1994a, 1994b), is very useful as an electrophysiological marker of attentional object selection.

A recent surge of interest in this ERP component enabled researchers to describe in detail the nature of the neuro-cognitive process it is reflecting. The N2pc component can be observed in response to targets defined by single features (e.g, colour, orientation, motion, or shape; Girelli & Luck, 1997; Kiss, Van Velzen, & Eimer, 2008), by a conjunction of features (Luck, Girelli, McDermott, & Ford, 1997) as well as in response to distractors which share target-defining features (Eimer et al., 2009). Importantly, this component seems to reflect a process of selection of *objects* in external space and not of spatial locations per se. Woodman, Arita, and Luck (2009) demonstrated that the N2pc component can be triggered by a placeholder indicating the likely cued location of an upcoming target, but not by an empty cued location. This finding is in line with the results of a MEG study (Hopf et al., 2000), which localised the neural generators of N2pc in the occipitotemporal cortex, known to be involved in object identification (Goodale & Milner, 1992). Of note, early parts of the N2pc component might be driven by neural activity arising in PPC (Hopf et al., 2000).

The studies described above strongly suggest that the N2pc component reflects an attentional mechanism that is qualitatively different from the attentional mechanisms described in the previous sections. As previously discussed, the P1 component reflects an early selection bias for stimuli (target or nontarget) presented at attended locations, which is likely to occur during the parallel feedforward 'sweep' of visual processing (Lamme & Roelfsema, 2000). In contrast, the N2pc component is typically thought to be triggered on the point of detection of target-defining features in the course of this initial object analysis, which triggers a spatially selective perceptual processing bias through feedback signals from higher-level top-down control areas, such as PPC. Moreover, N2pc is regarded as a correlate of the initial selection of the target object among distractors, i.e., a process that precedes the subsequent detailed analysis of specific object features but follows shifting attention to a target-candidate location. No N2pc is observed in cases where attention is shifted to the

likely location of an upcoming visual target (Kiss et al., 2008), demonstrating that the process which the N2pc component reflects is not present at this preparatory stage (but see Kiss et al., 2013, for a recent alternative for the mechanism underlying the N2pc component). Additionally, N2pc triggered by a target object in the search array is not modulated by the difficulty of the task to be performed on the target (i.e, target localisation vs. categorisation of the target diamond as left- or right-ward pointing; Mazza, Turatto, Umiltà, & Eimer, 2007).

Another issue concerning the N2pc component is whether the component reflects the target selection or, instead, the distractor inhibition. The latter account, according to which N2pc is an index of spatially-specific distractor suppression activated when target and distractors compete for selection, was promoted by early studies in this area: The component was shown to be present only in contexts where targets were surrounded by task-irrelevant distractors and its amplitude to be enhanced by increasing number of distractors that need to be suppressed (Luck et al., 1997). However, other studies demonstrated that N2pc is triggered by a target object even when the target is accompanied by a single distractor, and that N2pc amplitude is not affected by spatial proximity between the two stimuli (Eimer, 1996; Mazza, Turatto, & Caramazza, 2009), thus supporting interpretation of the N2pc component in terms of enhancement of spatially selective perceptual processing of target objects. An arising consensus in this debate is that the N2pc component is in fact composed of two subcomponents, where one is reflecting target processing and the other is a correlate of distractor inhibition. Several ERP studies (Eimer & Kiss, 2008; Hickey, Lollo, & McDonald, 2008; Sawaki & Luck, 2010) have recently demonstrated that distractors can trigger a lateralised component which resembles N2pc in its latency and topography. This component has been termed ‘a contralateral positivity’, or ‘distractor positivity’ (Pd; Hickey et al., 2008), because, in contrast to N2pc, its polarity is positive. The notion that Pd reflects active top-down distractor inhibition was supported by findings demonstrating its presence solely in contexts of high demands for attentional selectivity, e.g, where a salient colour singleton distractor appeared in a search array composed of nonsalient target letters (Sawaki & Luck, 2010).

The N2pc component was particularly useful in resolving the long-standing debate on the relative importance of top-down and bottom-up factors in the control of attentional selectivity (see Section 1.1.2 for more details). In an attempt to reconcile the seemingly contradictory findings from behavioural studies employing different paradigms, Kiss, Grubert, Petersen, and Eimer (2012) investigated the effect of changing temporal task demands on the ability of salient task-irrelevant feature singletons to capture visual

attention. They presented participants with search arrays containing a shape singleton target and a colour singleton distractor while manipulating the duration of the display, i.e., it could last until response execution (cf., Theeuwes, 1991) or for only 200 ms (cf., Folk et al., 1992). In the long-exposure condition, the colour distractors elicited a reliable N2pc, replicating the previous findings of Hickey, McDonald, and Theeuwes (2006) that salient distractors can capture attention even in cases where they are task-irrelevant. In contrast, in the short-exposure condition, a Pd, but no N2pc, was observed in response to the colour singleton. In other words, in circumstances in which selection of salient distractors would severely disrupt performance (e.g., in short-duration search arrays where there is limited time to reorient attention to the target), attentional capture by irrelevant stimuli was prevented by means of top-down inhibition mechanisms. Additionally, the N2pc component provided evidence that in contexts requiring high levels of attentional selectivity even large differences in visual salience might have only limited influence on the ability of cues sharing target-defining features to capture attention. For example, Eimer and colleagues (2009) instructed participants to search for a specific feature value (e.g., red bars) in a target display which could be preceded by target-colour cues presented among five differently coloured items (low salience condition) or as singletons (high salience condition). N2pc components of strikingly similar properties were triggered by both types of cue, thus supporting the dominating role of top-down, goal-based mechanisms over bottom-up, salience-based mechanisms in the control of object selection in multi-stimulus contexts.

These recent N2pc findings suggest that in tasks requiring high levels of selectivity salient but irrelevant visual distractors are actively inhibited, while distractors with task-relevant features are automatically selected (cf., Folk et al., 1992). However, this distinction is less clear in contexts in which targets are defined as conjunctions of multiple features. In such contexts, a distractor that matches one of the features of a feature-conjunction target might automatically capture attention because it possesses the target-like attribute, but it also might be suppressed as a nontarget stimulus. These alternatives point to one of the most important questions concerning the control of spatial selection in real-life environments, where targets are frequently defined by a combination of features: Is attention during search for multi-feature target objects guided by integrated search templates or by separately represented target features? The N2pc component was critical in a recent demonstration that both accounts are correct. Kiss and colleagues (2013) instructed participants to search for target singleton bars that were defined by a specific combination of colour and size (e.g., red small bars). The target arrays were preceded by cue arrays that contained a spatially uninformative colour/size singleton that could have both, one, or neither of the two visual

target features. The singleton cues that matched only one of these two target features failed to trigger reliable RTs spatial cueing effects, indicative of a lack of attentional capture, and supportive of selection guidance by integrated object representations during conjunction search. Critically, the same partially target-matching singleton cues triggered reliable N2pc components, consistent with attention guidance by independent features. To reconcile these findings, Kiss and colleagues (2013) proposed that during search for multi-feature targets, all objects matching one of the target features initially trigger attentional capture, reflected by the N2pc component, but attention is then rapidly disengaged from the locations of only partially target-matching stimuli, resulting in the lack of behavioural spatial cueing effects for such cues.

The studies discussed in this section provide mounting evidence for the usefulness of the N2pc component in elucidating the relative importance of bottom-up and top-down factors in the control of selection of objects in space. As a marker of the attentional selection of stimuli that does not require an overt response, the N2pc component should be valuable in investigating how spatial selection of irrelevant objects can be biased by multisensory integration. First, by measuring electrophysiological responses to visual distractors paired with task-irrelevant tones, N2pc can provide insight into whether multisensory integration can enhance early selection of visual objects in a manner that is independent of top-down attentional control settings (see Experiments 6 and 7 in this thesis). Second, the N2pc component can reveal whether selection is controlled by integrated or separate representations of task-relevant features in contexts where observers are searching for targets defined by combinations of features from different modalities (see Experiments 8 through 11 in this thesis). Notably, as an ERP component that reflects spatially selective attentional processing in modality-specific visual areas, the N2pc component can provide a direct insight into whether multisensory integration can serve as a mechanism of bottom-up as well as top-down control of spatial selection of objects at early stages of selective cortical processing. Thus, due to the importance of the insights that the N2pc component can reveal about the interactions between multisensory integration and object selection, the N2pc component will be the focus of a large part of this thesis.

1.5. The present thesis

As described in Section 1.3, despite a surge in interest in the interactions between multisensory integration and visuo-spatial attention, several important questions still remain unanswered in respect to the specific mechanisms by which spatial selection can be biased towards bimodal objects in multi-stimulus contexts. Two fundamental forms of such bias were addressed in the present thesis, and a combination of behavioural and electrophysiological measures was employed to provide a comprehensive picture on the neuro-cognitive mechanisms that are the source of these biases. The first set of experiments addressed the question whether audiovisual synchrony can create a bottom-up bias in visual selection by increasing salience of visual objects paired with task-irrelevant tones. The second set of experiments focused on whether during the search for objects defined by a conjunction of visual and auditory features, attentional selection can be controlled in a top-down manner by integrated audiovisual search templates. With one exception, the experiments reported in the following chapters were conducted using the cueing paradigm designed by Folk and colleagues (1992), which has proved useful in research into the role of purely visual bottom-up and top-down mechanisms that are involved in the control of visuo-spatial attention in multi-stimulus contexts. Sections 1.5.1 and 1.5.2 present the aims and rationale of the experiments included in this thesis.

1.5.1. Mechanisms underlying a salience-based bias in visual selection towards synchronous audiovisual objects

As highlighted in Section 1.1.1, temporal coincidence between a visual object and a task-irrelevant, spatially uninformative tone can increase the bottom-up salience of this visual object, which, in turn, should enhance its ability to attract shifts of involuntary visual attention in multi-stimulus contexts. The studies discussed in Section 1.3.2 provided initial evidence that audiovisual synchrony can create a bias in visual selection, both in sequential and in spatial tasks. However, the multisensory enhancement effects reported in these experiments were shown to be heavily dependent on top-down cross-modal attention, which left unanswered two critical questions concerning the role of audiovisual synchrony as a source of salience-based bias in visual selection towards bimodal objects. If multisensory integration enhances the bottom-up salience of visual objects, does it increase the ability of visual objects to capture attention in all contexts in which multiple simultaneous events

compete for selection? Critically, is this form of attentional selection bias towards bimodal objects contingent on goals of the observer, or is it a genuinely bottom-up phenomenon?

To answer these questions, the spatial cueing paradigm developed originally by Folk and colleagues (1992, see Sections 1.1.2.4 and 1.4.3.3. for more details) was adapted to audiovisual contexts. Spatial cueing effects were measured as a behavioural index of the ability of irrelevant colour-change cues to capture attention. The critical comparison was between cue-induced attentional capture effects on trials in which these cues were presented concurrently with task-irrelevant spatially uninformative auditory stimuli (audiovisual trials) and trials where cues appeared without synchronous auditory events (unimodal visual trials). The first question was addressed by assessing whether attentional capture effects triggered by colour cues are larger on audiovisual relative to visual trials in contexts where attention is controlled by bottom-up salience, and whether these enhancements can be explained by tone-induced alertness (Experiments 1 and 2). To address the second question, attentional capture by colour cues on visual versus audiovisual trials was assessed in contexts where participants were searching for targets defined by a specific colour, and the cues either matched or did not match this target colour (Experiment 3). Subsequently, the roles of top-down search mode and of the relative salience of colour cues in the modulation of multisensory enhancements of visual attention capture was investigated (Experiments 4 and 5, respectively). The last two experiments in this part of this thesis employed the posterior contralateral component known as the N2pc component (see Section 1.4.3.3) in order to examine whether audiovisual enhancements of attentional capture are also observable when a more direct electrophysiological marker of visual object selection is used (Experiments 6 and 7).

1.5.2. Mechanisms underlying top-down control of object selection by integrated audiovisual object templates

As indicated in Section 1.1.2, effective cognitive functioning is critically dependent on the ability to preferentially select objects and events that match one's current behavioural goals. Research carried out in the past twenty years has provided converging evidence (for more details, see Sections 1.1.2.4 and 1.4.3.3) for the contingency of attentional capture by irrelevant visual objects on whether these events possess currently task-relevant features. However, fewer studies have investigated whether separate features or integrated object templates control object selection in contexts where targets are defined by conjunctions of features (Kiss et al., 2013). Critically, even less is known about mechanisms of attentional

control supporting search for targets defined by features from different modalities. It is unclear whether the typically multi-sensory nature of objects in real-life environments resulted in creation of effective mechanisms to control search for multimodally defined target objects, or whether search for multimodal targets is guided by separate representations of task-relevant features in separate within-modal templates. Thus, the research reported in the second part of the present thesis focused on two major questions aimed at providing a better understanding of control of object selection by attentional templates in naturalistic multimodal contexts: Can object selection be controlled by fully integrated bimodal object representations in contexts where target objects are defined by conjunctions of visual and auditory features? Do integrated audiovisual templates control attention in all audiovisual search contexts or are there other important factors that modulate attentional guidance during search for bimodally defined targets?

To address these questions, another variation of the Folk et al.'s (1992) cueing paradigm was employed, in which in different blocks participants searched for objects defined by a single visual feature or a conjunction of visual and auditory features. Spatial cueing effects and the N2pc components triggered by unimodal target-matching cues were used as behavioural and electrophysiological measures, respectively, of changes in the ability of such cues to capture visual attention as a function of unimodal versus bimodal task sets. To address the first question, the task-set contingent attentional capture by target-colour singleton cues was compared across search tasks in which targets were defined solely by colour (e.g., 'red bars') versus a conjunction of colour and pitch (e.g., 'red bars accompanied by high-pitch tones'; Experiment 8). In order to address the second question, it was assessed whether stronger effects of audiovisual templates on task-set contingent capture by target-matching cues can be induced by a knowledge-based mechanism in contexts where the target-defining pitch is highly predictive of presence of the bimodal target (Experiment 9). The last two experiments reported in this part of the Thesis assessed how the guidance of attentional selection by integrated audiovisual object representations is modulated by the relative salience of visual cues and by the intrinsic salience of the visual dimension on which audiovisual targets are defined (Experiment 10 and 11, respectively).

Chapter 2. Multisensory enhancement of attentional capture in visual search

The present chapter describes the first two of six experiments aimed at investigating whether multisensory integration can create a bottom-up bias in spatial selection towards visual objects which are paired with signals in other modalities. It seems intuitively correct to assume that our attention should be automatically drawn to those objects in multi-object contexts which are highly distinctive from their surroundings, and, hence, potentially important. As discussed in Section 1.1.2, studies in the visual domain, which employed various paradigms and methods, provided strong evidence that behavioural goals determine to which stimuli our spatial attention is involuntarily attracted. Visual objects that are salient but not relevant to the currently performed task typically are not preferentially selected in space despite their salience because their processing will be actively inhibited by top-down attentional control (for more details, see Sections 1.2.2 and 1.4.3.3). Notably, there is mounting research to suggest that multisensory integration can increase bottom-up salience of visual objects and, as a consequence, enhance their ability to attract involuntary shifts of attention (see Section 1.3.2.1). While this evidence suggests preferential attentional processing of synchronous events, in real life we rarely orient to single events in empty space. Rather, we are confronted with multiple simultaneous stimuli and are therefore posed with a challenge of how to select objects which are important for the task-at-hand. In this context, a fundamental question, which still remains open (see Section 1.3.3.2), is whether multisensory integration can provide visual objects paired with non-visual signals with an advantage during competition for selection with other stimuli by increasing their bottom-up salience. Research into the role of multisensory integration in the bottom-up control of selection of visual objects will provide a more ecologically valid and novel perspective on attentional control by showing how it operates across modalities.

When a visual stimulus temporally coincides with an event entering another modality, they often (see Talsma et al., 2010) become effortlessly combined together at very early, pre-attentive stages of information processing, creating an emergent salient multimodal stimulus. In support of some forms of multisensory integration taking place during the initial feedforward 'sweep' of cortical processing, synchronous bimodal events were shown to elicit neural responses larger than unimodal events at low-level 'sensory-

specific' cortices and, critically, at very short latencies (for reviews, see Driver & Noesselt, 2008; Kayser & Logothetis, 2007). Lakatos et al. (2007) proposed the likely neural mechanism underlying multisensory enhancement in the early-level cortices. By analysing current source density and multi-unit activity in the A1 of awake macaques, they found that a concurrent somatosensory signal enhanced neural responses to an auditory event only when the two signals were synchronous or when the latter followed the former after time intervals corresponding to gamma, theta or delta band oscillations that form the spontaneous A1 activity. Lakatos et al. (2007) concluded that in sensory-specific cortices bimodal synchrony enhances neural responses not by increasing the firing rates (i.e., in this instance multisensory integration does not drive activity in neurons over their action-potential thresholds), but by resetting the phase of the ongoing slow-wave activity in specific cortices, which ensures that the target signal is processed when the phase of the oscillatory activity is at its maximal excitability (i.e., multisensory integration modulates the neural response to the unimodal input). These neural findings are in line with the behavioural evidence that demonstrated that temporal coincidence of a primary stimulus with a task-irrelevant event in another modality strengthens its perceptual representation, increasing its subjective brightness and improving its detection. Importantly, if the visual event is presented centrally, spatial alignment between the two signals is not necessary for integration and perceptual benefits (see Koelewijn et al., 2010, for a review).

Considering the findings discussed above, it can be assumed that if audiovisual synchrony creates salient integrated objects, such objects should have a competitive advantage during competition for selection with unimodal visual objects (Desimone & Duncan, 1995). Existing literature (van der Burg et al., 2008a, 2008b) provides only initial evidence that selective visual selection can be biased in a bottom-up fashion towards objects accompanied by task-irrelevant signals from other modalities. Experiment 1 in this thesis provided more direct evidence for this salience-based selection bias by showing that, in context where selection is guided by local features contrasts (singleton-detection mode; Bacon & Egeth, 1994), the ability of irrelevant visual distractors to attract rapid involuntary shifts of attention and affect selection of subsequent target events is enhanced when these distractors are paired with task-irrelevant spatially diffuse tones. Results of Experiment 2 strengthened the interpretations of the observed attentional capture enhancements in terms of multisensory integration increasing the bottom-up salience of visual objects by evidencing that these enhancements could not be explained by tone-induced alertness.

Experiment 1. Multisensory enhancement of visual attentional capture in visual search

Introduction

As described in Section 1.1.2, at every point of time in real-life environments multiple objects are competing for processing resources. The role of attention in such situations is to resolve the competition, either in a bottom-up manner in favour of objects highly distinctive from their surroundings, or in a top-down manner in favour of stimuli relevant to the current goals. In their 'biased competition model' of visual attention, Desimone and Duncan (1995) argued that salient objects, such as feature singletons, are preferentially processed at multiple levels of the respective cortical hierarchy, what increases their chances of becoming selected and gaining access to further information processing stages concerned with representation in short-term memory and response preparation.

If multisensory integration can increase perceptual salience of a unimodal object, then pairing visual objects with task-irrelevant tones should enhance their ability to attract involuntary shifts of visual attention in multi-stimulus contexts. As discussed in the Section 1.3.2.1, evidence for multisensory enhancement of attentional capture in vision was provided by studies employing paradigms with sequential stimuli presentation (Olivers and van der Burg, 2008; Vroomen & de Gelder, 2000). Using the 'attentional blink' paradigm, Olivers and van der Burg (2008) provided strong evidence that synchronous audiovisual objects can attract attention shifts more effectively than unimodal objects, even when there is no incentive to preferentially select audiovisual stimuli. Attentional blink describes a phenomenon where identification of a second visual target (T2) in a rapidly changing sequential array is typically impoverished when the stimulus is presented within approximately 500 ms from the onset of the first target (T1). Olivers and van der Burg (2008, Experiment 3) showed that this drop in accuracy is no longer observed when T2 is paired with an irrelevant tone. As tones in this particular experiment were paired predominantly with the distractors in the AB array, they could have not be treated by participants as temporal cues indicating onset of the T2 (for more details, see Section 1.3.2.1). Hence, the most likely explanation of the observed performance facilitation was that the bottom-up salience of T2 was increased due to audiovisual synchrony, which enhanced the ability of the visual target to attract involuntary attention and become selected.

However, as noted before, we rarely attend to single objects appearing in empty space. Rather, multiple simultaneous objects competing with each other at every point in time and selective attention acts to resolve this competition. If audiovisual synchrony creates salient integrated objects, it is possible that in situations of competition with unimodal objects selection would be biased in their favour due to their higher levels of salience. Initial evidence in support of this possibility was provided by two studies employing quite different paradigms, i.e., a visual search task and a temporal order judgement task (van der Burg et al., 2008a, 2008b). In their first study, van der Burg et al. (2008a) investigated whether serial and inefficient visual search can be improved when the visual targets are paired with task-irrelevant spatially uninformative tones. They instructed participants to search for vertical or horizontal target lines among 23, 35 or 47 oblique distractors (see Figure 2.1) that continuously changed colour. Critically, the target also changed colour (as the only element in the display at a given moment), and on some blocks this change was synchronised with onset of a tone. As shown in Figure 2.2 (left panel), in tone-absent blocks the search was slow, lasting several seconds on average, and inefficient, with steep search slopes. However, when the target was presented with a concurrent tone search times were strongly reduced and search slopes approached flat line.

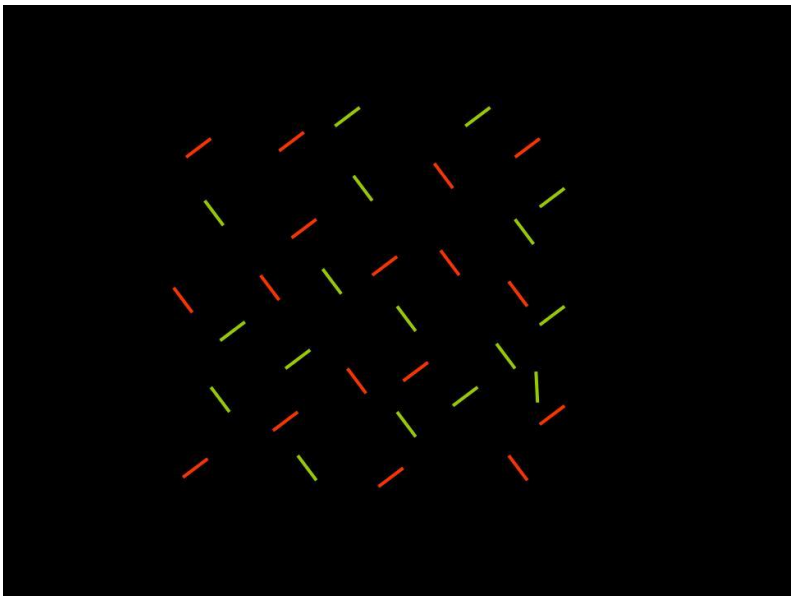

Figure 2.1. *Example of a 'pip-and-pop' visual search display, adapted from van der Burg et al. (2008a). In each trial both target and distractors changed colour continuously between a red and green colour on average every 900 ms and overall a colour change occurred every 50, 100 or 150 ms. Example depicts a target-present trial (i.e., green vertical line) with a set size of 36.*

While these results suggested that multisensory integration can aid resolution of a competition between multiple visual objects, several issues precluded interpretation of the findings strictly in terms of increased bottom-up salience of the synchronous target. The improvement of search observed on the tone-present blocks in Experiment 1 was shown in Experiment 4 to be strongly dependent on the co-occurrence of the tones with the target versus distractors. Namely, the magnitude of this benefit was directly related to the proportion of trials on which the tone was paired with the target colour change, rather than a colour change of one of the distractors preceding the target. This relation is depicted by flatter search slopes in Experiment 4a (Figure 2.2, middle panel), in which targets were paired with tones on 80% of all trials (and paired with distractors on the remaining 20% of all trials), compared to the search slopes in Experiment 4b, in which targets were paired with tones only on 20% all trials (and paired with distractors on 80% of all trials; Figure 2.2, right panel). These findings considered together with the results showing a lack of improvement following pairing targets with simultaneous irrelevant visual cues (i.e., brief offset of fixation point or flash of light behind the search display; Experiment 2) suggest that the tones facilitated search by being utilised by cross-modal attention as additional information about the onset of the target (for evidence in support of separate attentional resources, see Alais, Morrone, & Burr, 2006; Duncan, Martens, & Ward, 1997; Eimer, van Velzen, & Driver, 2002; Rees, Frith, & Lavie, 2001; Tellinghuisen & Nowak, 2003).

Initial evidence in support of at least partial contribution of increased bottom-up salience of the audiovisual target to search improvement was provided by van der Burg et al. (2008a) in their Experiment 5. In this experiment, targets always appeared alone, while the tones were presented with distractors, alone or absent on 40%, 40% and 20% of all trials, respectively. Overall, when the distractors that preceded the targets were audiovisual, larger search costs were observed compared to purely visual distractors, suggesting that tone presence enhanced the ability of irrelevant visual objects to attract involuntary shifts of attention. However, when the stimulus-onset asynchrony (SOA) between distractors and targets was 200 ms (i.e., when attention shifts to distractors were highly detrimental to target processing), bimodal distractors led to shorter search times and less steep search slopes than the visual ones, what contrasts with the expected automaticity of multisensory enhancement of involuntary attentional capture based on increased bottom-up salience of visual objects paired with diffuse non-visual signals.

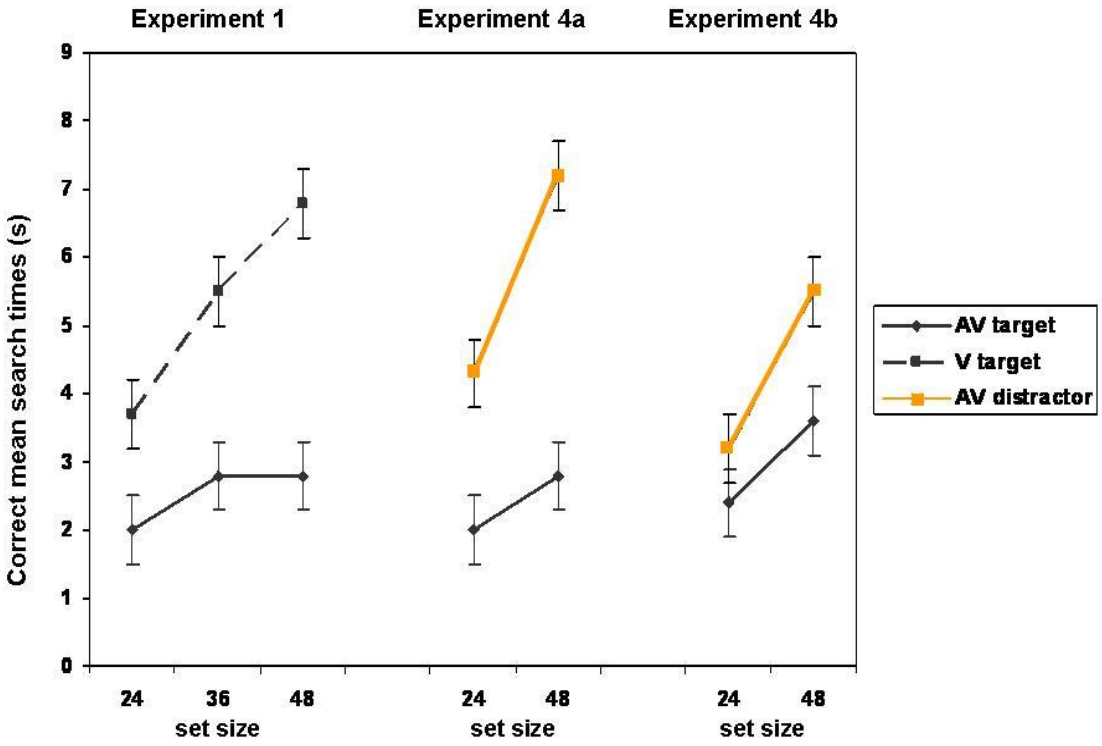

Figure 2.2. *Visualisation of the results from the 'pip-and-pop' study of van der Burg et al. (2008a). For Experiment 1, mean correct search times are shown as a function of set size (3 levels) and tone presence. For Experiments 4a and 4b, mean correct RTs are presented as a function of set size (2 levels) and synchronised item type. In Experiment 1 tone-present and tone-absent trials were presented in separate blocks, while in Experiment 4a and 4b different trial types were randomly intermixed within blocks.*

To provide more direct evidence for the role of audiovisual synchrony as a mechanism of bottom-up control of visual selection, van der Burg et al. (2008b) employed a temporal order judgement paradigm. In this study, participants were presented with two near-simultaneous lateral dots and instructed to judge which appeared first (a temporal order judgement, TOJ; Experiments 1 and 2) or detect their synchrony (a simultaneity judgement, SJ; Experiment 3). Next to each dot, nine irrelevant distractors constantly changed colour, and on half of all trials a spatially diffuse tone was presented together with colour change of one of the distractors on either side prior to the appearance of the first dot. Van der Burg et

al. (2008b) assumed that if shifts of involuntary attention affect temporal judgements (Shore, Spence, & Klein, 2001), and multisensory integration can enhance these shifts, the ability of visual distractors to bias perception will be stronger when they are paired with irrelevant tones. Observed results confirmed these predictions. In Experiment 1, primacy judgements were biased towards the side of a colour-change distractor only when it was paired with a tone. However, the influence of a response bias (i.e., a tendency to judge as first this of two events that was audiovisual; see also Jaśkowski, 1993) could not be ruled out in this experiment. When SJs, a less biased measure of temporal perception, were used (Experiment 3), two dots were perceived as synchronous more frequently on trials on which bimodal distractor was on the side opposite to the side of the first dot. However, in this experiment, the luminance of the target dots was much lower than in Experiment 1, and colour-change distractors were always paired with tones. These methodological made it unclear whether there was any difference in the ability of unimodal and bimodal distractors to capture attention (cf., Santangelo et al., 2006, who showed no difference in the ability of two types of cues to capture attention when visual stimuli were effective as exogenous cues).

Overall, findings of van der Burg and colleagues (2008a, 2008b) provided initial evidence that, by creating a bottom-up bias towards visual objects paired with non-visual signals, audiovisual synchrony can affect competition among multiple visual objects. However, mixed results (i.e., larger attentional capture by audiovisual distractors than visual distractors in the majority, but not all, SOA conditions in van der Burg et al. [2008a; Experiment 5]) or methodological issues (i.e., a lack of a visual-distractor condition against which the magnitude of attentional capture by audiovisual distractors could be compared in van der Burg et al. [2008b; Experiment 3]) precluded clear-cut interpretations.

Experiment 1 was designed to directly address the role of multisensory integration in creating a bottom-up bias in selection of visual objects. Folk et al.'s (1992) cueing paradigm was adapted for a cross-modal context. Search arrays with a colour-defined target were always preceded by colour-change cues (see Figure 2.3) by a time interval of 200 ms. Critically, these cues were paired with spatially uninformative task-irrelevant tones on 50% all trials. The role of audiovisual synchrony as a mechanism controlling spatial selection was investigated by comparing the ability of colour cues to affect selection of a subsequent colour target bar across trials where these cues were presented alone and presented together with tones. As the visual cues were spatially uninformative (i.e., they did not provide information about location of the subsequent target), the involuntary shifts of visual attention could be investigated. Spatial cueing effects (i.e., faster RTs to targets at cued vs.

uncued locations) were measured as a behavioural index of rapid involuntary attentional capture elicited by visual cues on tone-present and tone-absent trials.

Importantly, the present paradigm was designed in a fashion that minimised the influences of cross-modal attention. First, the tones were not informative about the identity of the target. Second, tones were presented from a loudspeaker located at the top of the screen, what ensured that tones provided no additional information about the location of the visual cues they were paired with. Also, visual cues were 100% predictive about the onset of the subsequent target, what should have precluded the tones from being utilised as a source of temporal information. Critically, para-foveal presentation of the colour-change cues should enable them to be automatically integrated with the tones despite the tones not being spatially aligned with the cues (Stein et al., 1996). Such a design ensured that any multisensory enhancements of attentional capture by audiovisual versus visual are due to their increased physical distinctiveness and not redundant information about their location. Importantly, a search strategy that promotes selection based on bottom-up salience was encouraged in subjects. Target bars were colour singletons (i.e., they were presented against five uniformly grey distractor bars) of two possible colours that were presented randomly and equiprobably within each block. Colour-change cues preceding the targets matched one of the target colours or had a third, nontarget colour. Previous studies (e.g., Eimer & Kiss, 2010) revealed that in task contexts where the target is a singleton and its colour is not predictable, participants adopt a 'singleton-detection mode' (Bacon & Egeth, 1994), i.e., they search for the target by allocating their attention preferentially to the most physically distinctive item in the display. Adoption of such a salience-based search strategy is indicated by attentional capture, measured by reliable cueing effects, elicited by all cues, irrespective of their colour. Thus, the critical prediction in Experiment 1 was that if audiovisual synchrony can enhance the ability of irrelevant visual objects to capture attention involuntarily by increasing their bottom-up salience, in contexts where object selection is guided by local feature contrast colour-change cues should elicit larger spatial cueing effects on trials on which they are paired with tones.

Method

Participants

Twenty paid volunteers (age range 19–31 years, mean age 24.7 years; 2 right-handed; 9 males) took part in the study. All had normal or corrected-to-normal vision and gave informed consent to participate in the study.

Stimuli, procedure, and analysis

The experiment was conducted in a dimly lit, air-conditioned room. Participants were seated at a distance of 100 cm from a 17-inch CRT monitor (75-Hz refresh rate). On each trial, a search array containing a target (50 ms duration) was preceded by a colour-change cue (i.e., one item in the 450-ms long base array changing colour in the cue array for 50 ms) which could be accompanied by a simultaneous tone (see Figure 2.3). All arrays, i.e., base array, cue array and search array, were composed of six elements presented equidistantly along the circumference of an imaginary circle at a distance of 2.1° of visual angle from the central fixation point. All visual stimuli were approximately equiluminant (~ 10.5 cd/m^2) and presented against a black background.

The search array consisted of five grey bars and one colour singleton target bar, each subtending 0.7° x 0.3°. For each participant, the target bar was equally likely to have one of two possible pre-defined colours, i.e., green and blue, blue and red, or red and green (CIE x/y chromaticity coordinates .285/.591 for green, .161/.128 for blue, and .621/.343 for red). Target colour sets were counterbalanced across participants. Target-colour bars were presented with equal probability and in a random order in one of four lateral locations, but never in the top or bottom location in the array (see Figure 2.3). The orientation of each of six bars was randomly determined on each trial. Participants were instructed to detect the target bar and judge its orientation (vertical vs. horizontal) by pressing one of two vertically aligned response keys (upper key for vertical, lower key for horizontal) with their left and right index fingers. Key-hand assignment was reversed after half of all blocks. Participants were instructed to respond as quickly and accurately as possible. In order to present colour cues that could be synchronised with tones, a base array was presented at the beginning of each trial and replaced immediately by the cue array (see Figure 2.3). Each base array consisted of six sets of four closely aligned grey dots (each subtending 0.1° x 0.1° visual angle; CIE x/y chromaticity coordinates .308/.345). In the cue array, one of the lateral sets was green, blue or red, resulting in a colour change. For each participant, two of the cue colours matched the two possible target colours (target-colour cues), while the third colour did not (nontarget-colour cues). In other words, for subjects searching for green and blue

targets, target-colour cues were green or blue, and nontarget-colour cues red. The base and cue arrays were identical, with the sole exception that one of the lateral sets in the cue array were of target or nontarget colour. Target- and nontarget-colour cues were presented with equal probability and in a random order within a block, and were spatially uninformative in regard to the location of the subsequent target bar.

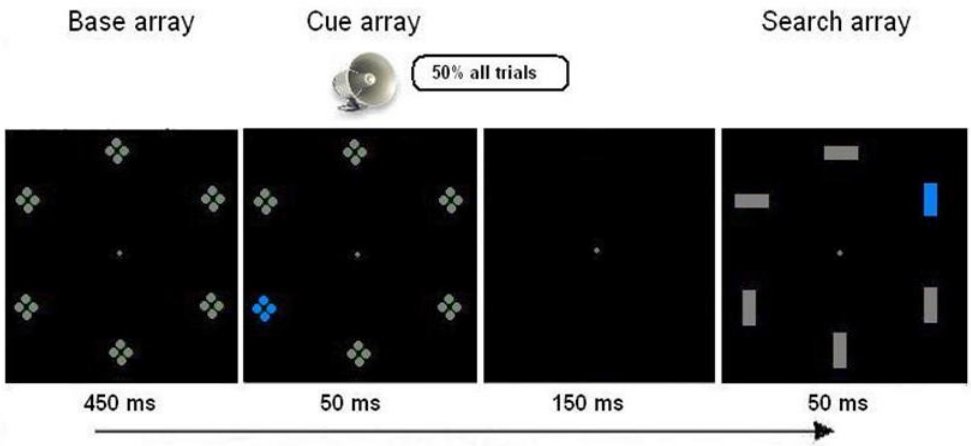

Figure 2.3. *The stimulus setup and trial sequence used in Experiment 1. Base array was immediately followed by a cue array which was followed by search array after an interval of 150 ms. Cues were presented concurrently with a tone on 50% of all trials. Colour change cues and targets were colour singletons. The example depicts a trial on which a target-colour cue invalidly indicated location of a blue target bar.*

Crucially, on half of all trials presentation of a colour-change cue in the cue array was synchronised with the onset of a pure-sine wave tone (2000-Hz frequency, 65 dB SPL intensity as measured around participant's head; 22.1 kHz sample rate, 8 bit, stereo) which was presented for 50 ms from a loudspeaker located at the top of the monitor. Tone-present and tone-absent trials were randomly intermixed within a block. The inter-stimulus interval between the offset of the cue array and onset of the search array was 150 ms, and the inter-trial interval was 1,500 ms. Participants completed 12 experimental blocks with 48 trials each, resulting in a total of 576 trials. RTs and error rates were analysed separately using a three-way analysis of variance (ANOVA) with cue type (target-colour cue vs. nontarget-

colour cue), tone presence (tone present vs. tone absent), and spatial cueing (target at cued vs. one of three uncued locations) as within-subjects factors.

Results

Premature and slow responses, defined as mean latencies shorter than 200 ms and longer than 1000 ms, respectively, were excluded from analyses of RTs and error rates, what resulted in a loss of less than 1% of all trials. The proportion of trials on which participants missed to respond was also below 1%. Figure 2.4 depicts RTs (line graphs) and error rates (bar graphs) for targets at cued and uncued locations, presented on tone-present and tone-absent trials, separately for target-colour and nontarget-colour cues.

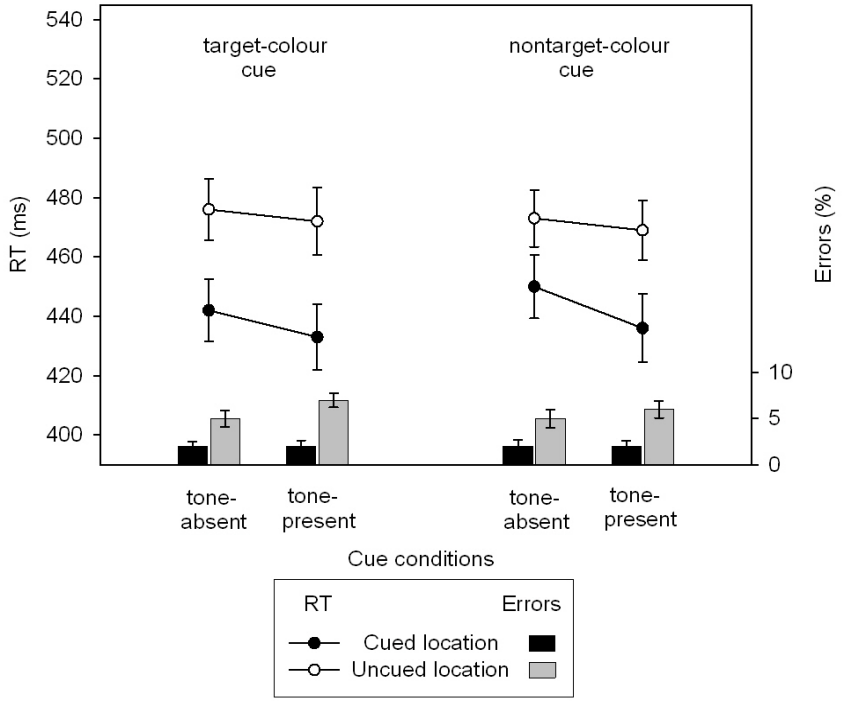

Figure 2.4. *Mean reaction times (RTs; line graphs) and error rates (bar graphs) in Experiment 1 measured in response to targets at cued and uncued locations, shown separately for target-colour and nontarget-colour cues and tone-present and tone-absent trials.*

Analysis of data from trials with correct responses revealed shorter RTs on tone-present than on tone-absent trials (452 ms vs. 460 ms), evidenced by a main effect of tone presence, $F(1,19) = 13.84$, $p < .001$, $\eta_p^2 = .42$. Overall, responses were also shorter to targets

presented at cued as opposed to uncued trials (440 ms vs. 472 ms), resulting in a main effect of spatial cueing, $F(1,19) = 172.\ 86$, $p < .001$, $\eta_p^2 = .91$. There was no two-way interaction between cue type and spatial cueing, $F(1,19) = 2.86$, $p = .11$, what suggested the fact that spatial cueing effects of comparable size were elicited by target-colour and nontarget-colour cues (see Figure 2.4). There was no main effect of cue type or a cue type x tone presence interaction, both F's < 1. Most importantly, a two-way interaction between tone presence and spatial cueing was observed, $F(1,19) = 4.71$, $p < .05$, $\eta_p^2 = .2$. This provided evidence that cueing effects elicited on tone-present trials (36 ms; $F(1,19) = 148.09$, $p < .001$), were reliably larger than cueing effects on tone-absent trials (29 ms; $F(1,19) = 93.27$, $p < .001$; see Figure 2.4). A lack of a three-way interaction between cue colour, tone presence and spatial cueing, $F < 1$, demonstrated that audiovisual enhancements of spatial cueing effects were not modulated by whether the visual cue shared the target colour or not.

As visible in Figure 2.4, errors were more frequent on tone-present than tone-absent trials (4.2% vs. 3.4%), evidenced by a main effect of tone presence, $F(1,19) = 5.49$, $p < .05$, $\eta_p^2 = .22$. Participants made more errors also on trials on which targets were presented at uncued, as opposed to cued locations (5.7% vs. 1.9%), which was reflected by a main effect of spatial cueing, $F(1,19) = 57.62$, $p < .001$, $\eta_p^2 = .75$. Additionally, the difference in frequency of errors between uncued and cued trials was larger on tone-present trials, resulting in a two-way tone presence x spatial cueing interaction, $F(1,19) = 4.79$, $p < .05$, $\eta_p^2 = .2$. None of the other effects was significant, all F's < 1.

Discussion

The aim of Experiment 1 was to provide evidence that multisensory integration can enhance the ability of irrelevant visual objects to capture attention in multi-stimulus contexts in which visual selection is guided towards objects of the highest physical distinctiveness. RT cueing effects elicited by colour-change cues were larger on trials in which these cues were paired with task-irrelevant spatially uninformative tones compared to trials in which the cues were presented separately. These results can be explained in terms of multisensory integration creating a bias in spatial selection by increasing bottom-up salience of visual objects: On trials on which task-irrelevant unimodal distractors were paired with spatially diffuse tones, the two signals became automatically combined together into a salient emergent bimodal object. Increased bottom-up salience enhanced the ability of such audiovisual distractors to attract involuntary shifts of attention, what had a direct effect on the speed of selection of subsequent targets in the search arrays.

The present findings are in line with a plethora of behavioural studies which demonstrated that multisensory integration can increase bottom-up salience of visual objects (e.g., Frasinetti et al., 2002; Noesselt et al., 2008; Stein et al., 1996) and enhance their ability of to capture attention (e.g., Olivers & van der Burg, 2008). Critically, Experiment 1 provided direct evidence that also competition among multiple visual objects can be biased towards synchronised audiovisual objects due to their increased bottom-up salience. The design used in this experiment minimised the possible influences of cross-modal temporal and spatial attention (cf., van der Burg et al., 2008a), which supports the explanation of the enlarged cueing effects on tone-present versus tone-absent trials in terms of a bottom-up bias in visual selection in the favour of synchronous audiovisual distractors. Also, spatial cueing effects elicited by visual and audiovisual cues were directly compared in Experiment 1, thus demonstrating directly that temporal co-occurrence of signals from different modalities can reliably enhance the ability of irrelevant unimodal objects to capture attention involuntarily (cf., van der Burg et al., 2008b). Notably, the cueing effects elicited by the colour-changes cues presented alone were also reliable, what indicates that, at least in the current paradigm, audiovisual synchrony can further increase the ability to attract involuntary shits of attention in distractors which are already effective as exogenous cues, thus providing evidence for audiovisual synchrony to have a distinctive effect on attentional selection to the form of multisensory integration described by Santangelo and Spence (2007a).

The interpretation of the present results in terms of multisensory integration enhancing visual attention capture by increasing bottom-up salience of visual objects is also substantiated by a lack of a two-way interaction between cue type and spatial cueing. These results reflected the fact that all singletons captured attention, thus confirming that participants adopted the singleton-detection mode (Bacon & Egeth, 1994) in which attention is preferentially drawn to the most distinctive objects in the display. In other words, when participants were allocating their attention preferentially to stimuli with the highest levels of bottom-up salience, audiovisual synchrony further enhanced the ability of irrelevant unimodal objects to capture attention. However, a singleton-detection mode also indicates the possibility that participants might have found all colour cues equally relevant to the task-at-hand (i.e., the target could have been defined as ‘any colour singleton’). This limits the conclusions that can be drawn from Experiment 1 about multisensory integration as a mechanism modulating attentional capture by visual objects irrespective of their task relevance. Experiment 3 was designed to directly address this issue.

However, before a more direct test of the bottom-up nature of the observed multisensory enhancements of the attentional capture can be conducted, an alternative explanation of the present results must be addressed. In Experiment 1, responses were faster and errors more frequent on tone-present relative to tone-absent trials. These findings suggest that tones had an alerting effect on performance, evidenced by a trade-off between the speed and the accuracy of responding. Hence, it is possible that the enhanced attentional capture in response to colour-change cues paired with tones might have been not due to multisensory integration, but merely alerting properties of the tones (cf., Callejas, Lupiáñez, & Tudela, 2004). While numerous studies provided evidence in support of the notion that signals from different modalities are combined together as long as they are presented within approximately 100 ms (e.g., Olivers & van der Burg, 2008; van der Burg et al., 2008a, 2008b), a time interval between tone and the target-colour bars of 200 ms employed in Experiment 1 would have been optimal for tone-induced alerting effects to develop. In order to directly address this alternative explanation of the tone-induced enlargements of spatial cueing effects, Experiment 2 was designed.

Experiment 2. The role of alertness in tone-induced enhancements of visual attentional capture in visual search

Introduction

Experiment 1 demonstrated that in multi-stimulus contexts in which attention is allocated to the most salient events, spatially non-predictive colour cues elicit larger cueing effects in situations in which they are paired with spatially diffuse tones, indicative of an enhanced ability of colour cues to capture visual attention as a function of tone presence. As discussed in Sections 1.1.1 and 1.2.1, as long as stimuli from different modalities are presented in close temporal proximity, they should be automatically combined into an emergent salient object whose ability to attract involuntary shifts of attention should be larger than that of a unimodal object. The tone-induced enhancements of attentional capture found in Experiment 1 are in line with this account. However, it is also possible that these enhancements were driven by the alerting properties of the task-irrelevant sounds (cf., Callejas et al., 2004).

Alerting is one of the mechanisms by which attention improves performance (Posner & Petersen, 1990; Posner & Rothbart, 2007), and auditory stimuli are known to have strong alerting properties (see Hackley, 2009, for a review). Because signals which increase alertness do not provide any information about the incoming events, they are thought to act on the late, responses-related stages of information processing, where they increase the readiness to respond to incoming events (Posner & Petersen, 1990). In other words, states of high alertness are characterised by responses that are faster, but also more erroneous than those observed in states of low alertness (i.e., there is a trade-off between speed and accuracy of response). Consequently, it is possible that in experiments in which the facilitation of attentional selection of visual targets accompanied by tones was measured through reduction in search times, e.g., in van der Burg et al. (2008a), this effect was not driven by multisensory integration, but by alerting properties of the tones. However, there is ample evidence that, for example, the 'pip-and-pop' phenomenon (van der Burg et al., 2008a) cannot be fully explained by increased alertness. First, shorter search times on tone-present versus tone-absent trials were accompanied by flatter search slopes. Second, the reduction of search time was in the order of seconds, as opposed to several milliseconds

typically found for alerting effects. Finally, no search benefits were observed when tones preceded the targets by 150 ms (van der Burg et al., 2008a, Experiment 3). Overall, these results provided strong evidence against the idea that tone-induced enhancements of visual selection can be explained solely by alerting properties of the task-irrelevant sounds.

In Experiment 1 of this thesis, the enhancing effect of tone presence on visuo-spatial attention was investigated by measuring how selection of visual targets is affected in cases in which they are preceded by irrelevant visual objects paired with spatially diffuse tones. Existing literature shows that tone-induced alertness should have little to no influence on orienting of attention (Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Fan, McCandliss, Sommer, Raz, & Posner, 2002; Fernandez-Duque & Posner, 1997). Fernandez-Duque and Posner (1997) argued that neural networks supporting these two major functions of attention operate independently of each other when facilitating performance: While alertness increases the readiness to respond to all subsequent stimuli irrespective of their location in the visual field, orienting of attention benefits perceptual processing of only those stimuli that fall within the focus of the attentional 'spotlight' (Posner, 1980; see Section 1.1.2 for more details). In one of their studies, Fernandez-Duque and Posner (1997, Experiment 4) employed a visual discrimination task ('+' vs. 'x'), in which a spatially informative cue (80% valid) preceded the target by a varying time interval of 100 or 400 ms. Crucially, on some trials, a spatially-unspecific tone could be presented; 400 ms before the visual cue, concurrently with it, or instead of it. Validity effects (i.e., faster RTs on validly vs. invalidly cued trials) were not affected by presence of tone, thus suggesting that the effect of tone-induced alertness on visual performance is independent of the effect of attentional orienting (see also Fan et al., 2002, Fan et al., 2005).

Notably, spatial cues in the majority of the studies whose findings argued against an interaction between these two functions were spatially informative. In other words, the conclusions about interactions between alerting and visual attention that can be drawn from this research might be limited only to the endogenous visual attention, thus leaving open the question of whether tone-induced alertness was responsible for stronger orienting of the exogenous visual attention that was observed in Experiment 1. In fact, in contrast to the body of evidence indicating that alerting and voluntary attention are independent, a study conducted Callejas et al. (2004) suggested that tone-induced alertness can increase the ability of visual cues to attract shifts of involuntary attention. In their study, each trial began with the presentation of a fixation point for a variable duration, followed after 400 ms by a spatially uninformative cue (50% valid), and followed after 100 ms by the target. Critically, on trials in which the cues were preceded by 400 ms by a tone, the cueing effects elicited by

these cues were enlarged, in line with the idea that larger shifts of involuntary attention triggered by visual cues paired with spatially diffuse tones are driven by tone-induced alertness.

In Experiment 1 of the present thesis, the effect of audiovisual synchrony on visual selection was investigated by assessing whether the ability of spatially uninformative visual cues to capture attention is increased when these cues are paired with tones. Thus, it is possible that the enlargements of cueing effects observed here on tone-present compared to tone-absent trials were due to tone-induced alertness. In support of this, shorter RTs and more frequent errors were shown on trials on which tones were present, indicative of a speed-accuracy trade-off in responding. Experiment 2 was designed to directly address the possibility that the enhancements of attentional capture observed on tone-present versus tone-absent trials were driven by tone-induced alertness, as opposed to multisensory integration of temporally coincident colour cues and tones. Design of Experiment 2 was identical to the one used in Experiment 1, with the sole exception that the tones were now presented at the beginning of the trial, concurrently with the onset of the base array. With the resulting SOA of 450 ms between tones and cues, the alerting effects of tones should be the largest, while the probability of integration of two signals should be minimal. If cueing effects are now enlarged on trials on which colour cues are paired with tones, this would provide strong evidence for an alerting account of the enhancements of attentional capture observed in Experiment 1.

Method

Participants

Twelve paid volunteers (mean age 25.8 years, age range 23–31 years; 1 right-handed; 3 males) took part in the experiment. All had normal or corrected-to-normal acuity of vision and gave informed consent to participate in the study.

Stimuli, apparatus, and procedure

The stimuli and procedure were exactly the same as in Experiment 1, with the exception that now the tone was presented concurrently with the base array at the beginning of the trial, what made it precede the onset of the cue by 450 ms. Additionally, a digital monitor with a

100-Hz refresh rate was used and the speaker was now located centrally behind the screen (not on top of the screen, as in Experiment 1). Apart from the fact that now the tone preceded the cue presentation, the design of the experiment was identical to Experiment 1.

Results

Figure 2.5 depicts mean RTs for correct responses and error rates for targets at cued and uncued locations, separately for target-colour cues and nontarget-colour cues on trials with tones present and tones absent. Preliminary and slow responses (defined as, respectively, responses faster than 200 and responses slower than 1000 ms) were excluded from analyses, what resulted in a loss of less than 1% all trials. The proportion of trials on which participants missed to respond was also below 1%.

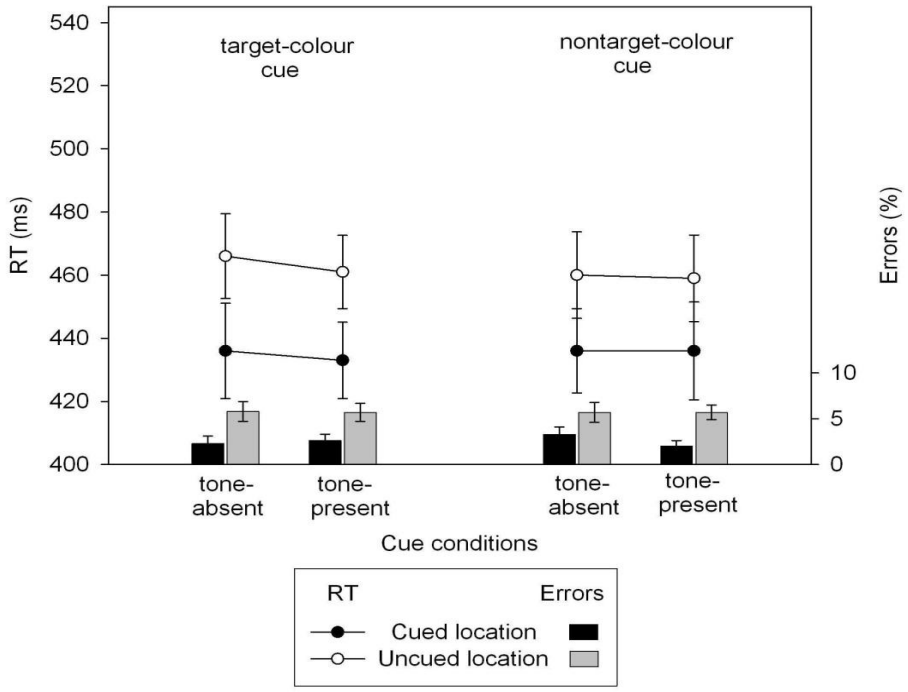

Figure 2.5. *Mean RTs (line graphs) and error rates (bar graphs) observed in Experiment 2 in response to targets at cued and uncued locations, shown separately for target-colour and nontarget-colour cues and tone-present and tone-absent trials.*

Responses to targets presented at the cued locations were faster than responses to targets presented at other locations (436 ms vs. 461 ms), what resulted in a main effect of spatial cueing, $F(1,11) = 31.03$, $p < .001$, $\eta_p^2 = 0.74$. This main effect was not further

modulated by type of cue, $F < 1$, which indicated that reliable spatial cueing effects were elicited again by all colour singleton cues. There was no main effect of cue or interaction of cue type and tone on RTs, both F's < 1. Somewhat surprisingly, there was only a tendency for faster responses on tone-present than tone-absent (448 ms vs. 450 ms) trials, as shown by a lack of main effect of tone presence, $F(1,11) = 1.79$, $p = .1$. Critically and in contrast with Experiment 1, no interaction between tone presence and spatial cueing was now observed, $F < 1$. This indicated that in Experiment 2 spatial cueing effects elicited by colour cues on tone-present trials (25 ms, $F(1,11) = 23.32$, $p < .001$) were not larger[1] than the cueing effects these cues elicited on tone-absent trials (27 ms, $F(1,11) = 39.46$, $p < .001$; see Figure 2.5). There was also no three-way interaction between cue type, tone presence and spatial cueing, $F < 1$. In other words, under conditions when the tones were presented 450 ms before the cue display, tone presence did not modulate the spatial cueing effects elicited by colour cues.

As visible in Figure 2.5, participants made fewer errors on trials in which the target location was previously cued relative to uncued locations (2.5% vs. 5.7%), as evidenced by a main effect of spatial cueing, $F(1,11) = 15.68$, $p < .01$, $\eta_p^2 = .59$. There was no main effect of cue, $F < 1$, and the interaction between cue type and tone presence was not significant, $F(1,11) = 1$, $p = .34$.A lack of main effect of tone presence, $F < 1$, suggested that errors were made with similar frequency on tone-present and on tone-absent trials. Crucially, there was no interaction between tone presence and spatial cueing, $F < 1$, and there was no three-way interaction involving cue type, $F < 1$.

Discussion

In order to investigate whether the enhancement of attentional capture by visual cues observed on tone-present relative to tone-absent trials in Experiment 1 could be explained by tone-induced alertness, in Experiment 2 the tones were presented concurrently with the onset of the base array (as opposed to the onset of the cue array, as in Experiment 1). With an SOA of 450 ms between the tone and the cue, automatic multisensory integration of the two signals should have been unlikely, while the alerting effects of tones on attentional selection of the visual targets should have been maximal. As predicted and in contrast to Experiment 1, no difference was observed in Experiment 2 between the spatial cueing effects that were elicited by colour-change cues on tone-absent and tone-present trials. This

[1] In fact, in Experiment 2, spatial cueing effects on tone-present trials were numerically smaller than on tone-absent trials.

pattern of results indicates that in Experiment 1 the presence of tones enhanced the ability of the irrelevant visual cues to capture attention because the temporal co-occurrence of the two signals led to their automatic multisensory integration, which increased bottom-up salience of the visual distractors.

The findings from Experiment 2 are in line with numerous studies which suggested that alerting and orienting have divergent and independent effects on visual selection (e.g., Fan et al., 2002, 2005). Notably, results from the present experiment are in contrast with the findings of Callejas et al. (2004), who showed that irrelevant visual cues elicited larger cueing effects when preceded by 400 ms by tones. Callejas et al. (2004) interpreted the enlargements of spatial cueing effects they observed on tone-present relative to tone-absent trials in terms of tone-induced alertness. However, explanation of their results in terms of multisensory integration is more likely. While the typical time window within which events from different modalities automatically interact with each other is 100 ms (see Holmes & Spence, 2005), it is a generally consensus that the possibility of integration decreases monotonically with a gradual increase of asynchrony between two events, rather than disappearing sharply (Meredith, Nemitz, & Stein, 1987; see also van der Burg et al., 2008a). Furthermore, according to the framework proposed by Talsma et al. (2010), low rates of stimulation within modalities, as in the study of Callejas et al. (2004), where tones and cues were the only non-stationary events presented before the target onset, likely increased the possibility of automatic integration of events occurring in close temporal proximity.

Interestingly, Experiment 2 provided indirect evidence in support of the explanation of the tone-induced enhancements observed by Callejas et al. (2004) in terms of multisensory integration. Namely, in spite of the fact that tones in Experiment 2 were presented at intervals that should have increased their alerting properties, their presence did not speed up the responses or increase the frequency of errors in this experiment. Thus, in contrast to Experiment 1, there was now no speed-accuracy trade-off on tone-present trials. A likely explanation for these findings is that temporal coincidence of the onsets of tones and of the base array led to their automatic integration, what in turn could have diminished the effect of alerting properties of the tone on visual selection: As base array was a homogenously coloured object spanning the whole area within which target search was to take place, the audiovisually induced increase of its bottom-up salience did not affect the attentional processing of the subsequent cues, or targets. Extrapolating to the study of Callejas et al. (2004), the tones and the visual cues were the only transient stimuli in the task they employed, what could have contributed to their automatic (see Section 1.1.1.2) or expectation-based (see ‘unity assumption’, Section 1.3.3). The critical difference between

Experiment 2 of this thesis and the study of Callejas et al. (2004) is that in the former case the tone fused with a display-wide visual object, while in the latter case the tone was bound with a visual spatial cue that indicated target location. Furthermore, the tone-induced enlargements of spatial cueing effects found by Callejas et al. (2004) were in the order of 10 ms, which is similar to the size of effect of tone presence on cueing effects found in Experiment 1 of the present thesis.

In summary, Experiment 2 demonstrated that in contexts where tones are presented at time intervals which should increase their alerting properties and decrease the likelihood of their integration with spatially uninformative visual cues, spatial cueing effects, indicative of attentional capture by visual cues, are no longer enlarged by tone presence. These findings are in line with the account according to which the tones were automatically integrated with another visual stimulus with which they were in the closest temporal proximity. Overall, the pattern of results found in Experiment 2 provides converging evidence that in cases where irrelevant visual objects are paired with irrelevant signals from other modalities, bottom-up salience of these objects is increased due to multisensory integration, which in turn enhances their ability to attract involuntary shifts of visual attention in multi-stimulus contexts.

General Discussion

Initial evidence that in situations where multiple visual objects compete for attentional selection multisensory integration can bias this competition in a bottom-up fashion towards synchronous audiovisual events was provided by van der Burg et al. (2008a, 2008b). In their studies, irrelevant visual distractors were shown to capture attention to a larger extent when paired with task-irrelevant spatially diffuse tones. However, mixed findings and methodological problems precluded treating of these enhancements as strong evidence in support of the idea that audiovisual synchrony can create a reliable bottom-up bias in selection of visual objects in multi-stimulus environments. This notion was directly tested in Experiment 1, in which the spatial cueing paradigm (Folk et al., 1992) was employed to assess the ability of irrelevant visual distractors to attract involuntary shifts of attention when they are paired with task-irrelevant spatially uninformative tones. Spatial cueing effects elicited by colour-change cues were enlarged on trials on which these cues were accompanied by tones, indicative of multisensory enhancement of rapid involuntary attentional capture. These results provided the first strong evidence that multisensory

integration can bias competition among multiple simultaneous visual objects in favour of the visual objects which coincide with irrelevant signals from other modalities. Crucially, the enhancements of attentional capture observed on tone-present versus tone-absent trials, indicative of such a bias, could not be explained by tone-induced alertness (Experiment 2).

Considering Experiments 1 and 2 together, the ability of irrelevant visual distractors to attract rapid involuntary shifts of attention is increased in cases where these distractors are presented concurrently with irrelevant uninformative tones, consistent with the existing neural and behavioural research that indicated that, by increasing the bottom-up salience of visual objects, multisensory integration can enhance visual selection (see Sections 1.2.1 and 1.3.2.1). The experiments described in Chapter 3 addressed the question whether the multisensory enhancement of visual object selection found in Experiment 1 is a genuine bottom-up phenomenon.

Chapter 3. Multisensory integration as a mechanism for creating a bottom-up bias in visual object selection

A recent but already substantial body of research provides strong support for the idea that temporal co-occurrence of visual and auditory events is often sufficient for integrated bimodal events to be created at low stages of the cortical hierarchy. Human ERP studies (e.g., Fort et al., 2002; Giard & Peronnet, 1999; Molholm et al., 2002) demonstrated that cross-modal interactions can occur as early as 50 ms after stimulus onset, which indicates that feedback projections from heteromodal convergence zones are unlikely to be their source. Instead, these interactions seem to be supported by thalamus-mediated feedforward or by direct lateral connections between primary and secondary visual and auditory cortices that were indicated by animal tracer studies (Cappe et al., 2009; Falchier et al., 2002). The timing of the synchrony-based multisensory integration and the type of neural projections that supports it suggest that audiovisual synchrony not only creates salient bimodal events that might be preferentially processed by visual selective attention, but also that the mechanism by which synchrony controls the visual selection does not depend on top-down feedback. Consistent with both these assumptions, previous behavioural studies demonstrated that visual events and objects which are temporally coincident with tones can be more easily perceived and attended to even in contexts where the tones provide no additional information about the identity, or location in time or space of the former (Noesselt et al., 2008; Olivers & van der Burg, 2008; Stein et al., 1996; Vroomen & de Gelder, 2000). However, as described in the Section 1.3.2, conclusions from those studies that aimed at investigating whether automatic multisensory integration can have a direct effect on spatial selection in vision by creating a bottom-up bias towards visual objects accompanied by non-visual signals are weakened by confounds in the design or by mixed results, and the first unequivocal evidence for existence of such effects was provided only very recently (Experiment 1 from Chapter 2 and Experiment 3 from the present chapter are reported in Matusz & Eimer, 2011).

Findings presented in the Chapter 2 demonstrated that in contexts where attention is deployed preferentially to the most salient events in the visual field (singleton-detection mode; Bacon & Egeth, 1994), colour cues can trigger larger capture effects when

accompanied by task-irrelevant tones (Experiment 1). As these results are unlikely to be explained by cross-modal endogenous attention or tone-induced alertness (Experiment 2), their interpretation in terms of multisensory integration creating a bottom-up bias in visual selection is plausible. However, because the multisensory enhancement of attentional capture was found in a context where participants were searching for the 'odd-one-out', it is possible that audiovisual synchrony triggers larger attention shifts to visual objects accompanied by tones via a mechanism that is contingent on whether these events share task-relevant features or not. Thus, the aim of the experiments described in Chapter 3 was to address directly the issue of whether the mechanism by which multisensory integration creates a bias in spatial selection of visual objects is purely salience-based in nature. For this purpose, in Experiment 3 a colour-specific feature search mode (e.g., 'target is a red bar'; Bacon & Egeth, 1994) was encouraged in participants. The rationale was that if enhanced spatial effects are still observed for both target-colour cues as well as non-target colour cues, this would provide strong support for the bottom-up nature of the audiovisually induced bias in attentional selection of objects in space. Surprisingly, in Experiment 3, no evidence of tone-induced modulation of spatial cueing effects was shown for either type of colour cues. Thus, Experiments 4 and 5 were conducted to identify the factors that determine the presence of audiovisual enhancement of spatial selection of visual objects. In line with the salience-based nature of the investigated mechanism, multisensory enhancement of visual attentional capture was modulated by factors pertaining to the physical distinctiveness of objects (Experiment 5), but not those related to task requirements (Experiment 4).

Experiment 3. The bottom-up nature of bias in visual object selection towards audiovisual objects

Introduction

The purpose of Experiment 3 was to directly assess whether multisensory integration enhances spatial selection of visual objects accompanied by non-visual signals by a mechanism that is modulated by top-down factors or whether this mechanism is genuinely bottom-up in nature. Simultaneity-induced enlargements of spatial cueing effects reported in the previous chapter in Experiment 1 cannot be treated as strong evidence in support of the bottom-up explanation, as the fact that participants adopted a singleton-detection mode in Experiment 1 opens the possibility that they regarded all types of colour cues as task-relevant. Unanimous evidence that multisensory integration enhances visual selection by a mechanism that is salience-based in nature can only be provided by task contexts in which simultaneity-induced enhancements of the ability of visual objects to capture attention are observed irrespective of whether these objects share features with the current target or not.

To directly address this issue, in Experiment 3 observers were encouraged to search for targets on the basis of a specific feature value (e.g., 'targets are bars of red colour'; feature-search mode, Bacon & Egeth, 1994). So far, this search strategy was used solely to investigate the role of purely visual bottom-up activation (i.e., local differences in contrast on hypothetical feature maps; Treisman & Gelade, 1980) in the control of attentional selection of visual objects in space. It is a well-established finding that in such high-selectivity contexts salient visual distractors that do not match target-defining features do not trigger involuntary shifts of attention (e.g., Eimer et al., 2009, 2010; Lamy et al., 2003, 2004; Lien et al., 2008), a phenomenon known as 'task-set contingent attentional capture' (Folk et al., 1992). In contrast, the modulations of selection of visual objects that are driven by audiovisual simultaneity would be expected to operate independently of a top-down colour task set, because multisensory integration should create a bias in spatial selection of visual objects by up-modulating the activation elicited by physical objects at stages of the cortical hierarchy that precede the stages at which attentional control is present (see Sections 1.2.1 and 1.4.2, for more details).

To discourage participants from searching for the 'odd-one-out' in Experiment 3, the target-colour bar in the search array was now surrounded by five differently coloured

distractor bars. As the targets were no longer feature singletons, participants were forced to adopt strategy of searching for a specific colour value (i.e., 'targets are bars of red colour'). Participants were also informed that they would be searching for bars of one pre-specified colour throughout the whole experiment. Additionally, the colour-change cues were now also presented against a background of five differently coloured items, in order to maintain similarity of cue and target displays that was present in Experiment 1. The aim of this manipulation was to prevent the capture effects that are triggered by colour cues, and their modulation by synchronous tones, from being artificially reduced by a difference in display-wide properties between cue and search displays (Gibson & Kelsey, 1998; Lamy et al., 2004, but see Eimer et al., 2009). Importantly, colour-change cues could now either match the target colour (target-colour cues), or have another, nontarget colour (nontarget-colour cues). That is, if a participant was instructed to look for a blue bar, on half of the trials search arrays would be preceded by target-matching blue colour-change cues, and by target-nonmatching red colour-change cues on the remaining trials. Similarly to Experiment 1, on half of all trials these cues were accompanied by task-irrelevant tones. In line with purely visual studies in which the feature-search mode was encouraged (e.g., Eimer et al., 2009; Eimer & Kiss, 2010), reliable spatial cueing effects were expected only for cues that matched the target colour. The critical question was whether the simultaneity-induced enlargements of spatial cueing effects would be dependent on the colour task set. If multisensory enhancement of attentional capture in visual search is contingent on attentional control settings, only cueing effects elicited by target-colour cues, but not by nontarget-colour cues, should be enlarged on tone-present trials. In contrast, if audiovisual synchrony enhances attentional capture by increasing the bottom-up salience of visual objects, larger cueing effects should be observed on tone-present trials versus tone-absent trials for both target-colour and nontarget-colour cues.

Method

Participants

Twenty-two volunteers took part in this experiment (age range 18–33 years; mean age 24.2 years; 1 left-handed; 8 males). All had normal or corrected-to-normal vision, gave informed consent to participate in the study.

Stimuli, procedure and design

Stimuli, experimental procedures and design were identical to the ones employed in Experiment 1, with a few exceptions. In order to maximise participants' incentive to adopt a colour-specific task set, the search array was now heterogeneous (see Figure 3.1). Each of the five distractor bars in the search display was now randomly assigned a different colour from a set of six task-irrelevant colours with different CIE chromaticity coordinates (purple .220/ .119; turquoise .248/.429; green .285/.591.; pink .493/.281; orange .558/ .387; yellow .432/.485).

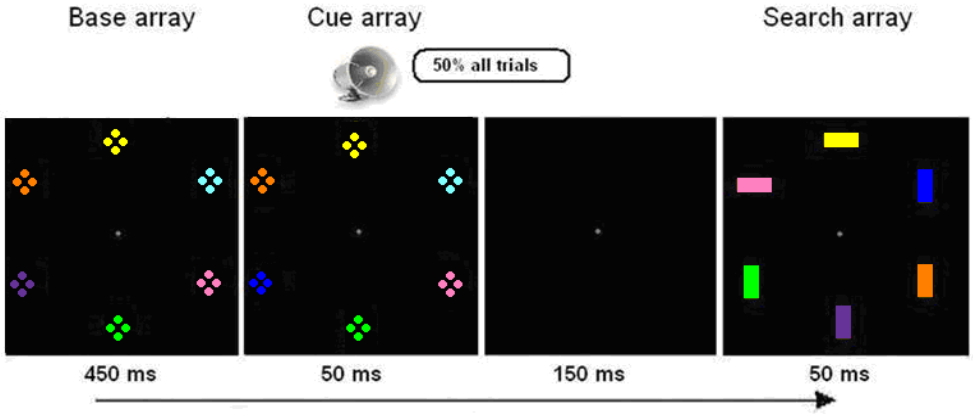

Figure 3.1. *The stimulus setup and trial sequence used in Experiment 3. Colour-change cues and targets were presented against a heterogeneous background of five differently coloured distractors. The example depicts a trial on which a target-colour cue invalidly indicated location of a blue target bar.*

To retain the similarity of the cue and the target arrays, the colour-change cues as well appeared against a background of five differently coloured items. Consequently, the base array was also heterogeneous, with each set of dots assigned randomly one of the six irrelevant colours from the same set used for the distractors in the search array (see Figure 3.1). Participants were now instructed to search for a bar of one predefined colour: For half of them the target bar was blue, for the other half the bar was red (with the order

counterbalanced across subjects). For each participant, the colour change in the cue display could be either to a target-bar colour matching colour (target-colour cue), or to another, non-target colour (nontarget-colour cue). The critical manipulation was that on half of all trials each of these colour changes was presented concurrently with a tone, which was identical to Experiment 1. Again, all the stimuli used were equiluminant (~10.5 cd/m^2). Participants completed 8 experimental blocks with 64 trials each, resulting in a total of 512 trials.

Results

Figure 3.2 shows mean RTs for correct responses and error rates for targets at cued and uncued locations, separately for target-colour cues and nontarget-colour cues on trials where tones were present and where tones were absent. Exclusion of data from trials on which responses were faster than 200 ms or slower than 1000 ms led to a loss of less than 1% of all trials. Participants missed to respond on less than 1% of all trials.

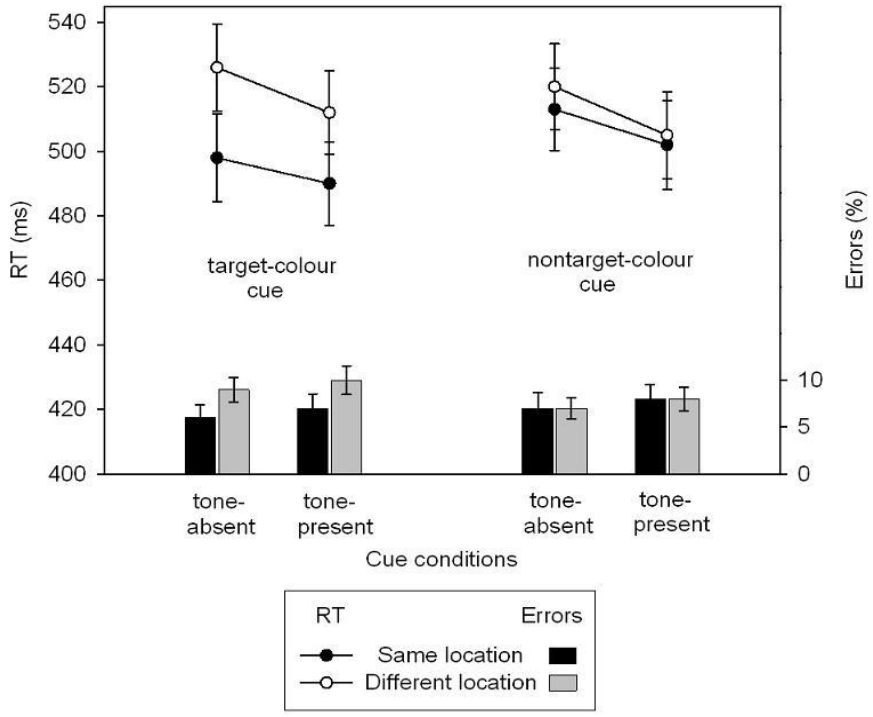

Figure 3.2. *Mean RTs (line graphs) and error rates (bar graphs) found in Experiment 3 in response to targets at cued and uncued locations, shown separately for target-colour and nontarget-colour cues and tone-present and tone-absent trials.*

Responses to targets presented at cued locations were faster than to targets presented at uncued locations (501 ms vs. 516 ms), as shown by a main effect of spatial cueing, $F(1,21) = 18.32$, $p < .001$, $\eta_p^2 = .47$. This effect was strongly modulated by the type of cue presented, $F(1,21) = 20.95$, $p < .001$, $\eta_p^2 = .5$, demonstrating the typical task-set contingent attentional capture. Whereas the visual cues that matched the target colour elicited reliably presented spatial cueing effects of 25 ms, $F(1,21) = 31.45$, $p < .001$, cueing effects elicited by nontarget-colour cues were not significantly different from 0 ms, $F(1,21) = 1.64$, $p = .21$ (see Figure 3.2). A main effect of tone presence, $F(1,21) = 16.31$, $p < .001$, $\eta_p^2 = .44$, indicated that responses were faster on trials on which targets were preceded by visual cues accompanied by tones, relative to trials on which visual cues were presented alone (502 vs. 514 ms). Surprisingly, tones now failed to enlarge the cueing effects elicited by colour cues[1], as suggested by a lack of an interaction between tone presence and spatial cueing, $F(1,21) = 2.47$, $p = .13$. There was also no three-way interaction between cue type, tone presence and spatial cueing, $F < 1$, indicating there was no evidence of the modulation of the effect of tone presence on spatial effects by the type of colour cue.

Participants made fewer errors on the trials on which the target location was cued than on the trials on which target appeared at uncued locations (6.9% vs. 8.5%), as shown by a main effect spatial cueing, $F(1,21) = 5.84$, $p < .05$, $\eta_p^2 = .22$. A cue type by spatial cueing interaction, $F(1,21) = 5.2$, $p < .05$, $\eta_p^2 = .2$, provided evidence that this effect was driven by more frequent errors on uncued versus cued trials for target-colour cues, $F(1,21) = 12.68$, $p < .01$, as opposed to nontarget-colour cues, $F < 1$ (see Figure 3.2). There was a statistically non-significant tendency for errors to be more frequent on tone-present relative to tone-absent trials (8.1% vs. 7.2%), $F(1,21) = 3.75$, $p = .066$, $\eta_p^2 = .15$. None of the remaining interactions was significant, with all F values < 1.

Discussion

A feature-search mode (Bacon & Egeth, 1994) was encouraged in Experiment 3 to investigate whether the multisensory enhancement of visual attention capture would be modulated by top-down task set. In line with participants adopting a colour-specific feature-search mode, reliable RT spatial cueing effects were now elicited only by the colour-change cues that matched the target colour (Eimer et al., 2009; Lamy et al., 2004). In a striking

[1] Interestingly, in Experiment 3 the direction of influence of tones on the spatial cueing effects was inverted, i.e., there was a tendency for the spatial cueing effects elicited by colour cues to be decreased on trials on which these cues were paired with tones (17 ms vs. 12 ms).

contrast to Experiment 1, no enlargement, but instead a tendency for a reduction of RT spatial cueing effects was observed on trials on which the colour cues were paired with task-irrelevant tones. These rather surprising findings suggested that audiovisual synchrony does not always lead to an enhancement of the ability of visual objects to attract attention shifts in multi-stimulus contexts.

There are two possible explanations for this pattern of results. According to the first account, in order for audiovisual synchrony to create a reliable bottom-up bias in spatial selection towards a visual object accompanied by a non-visual signal, the relative salience of the emergent bimodal object needs to reach a certain threshold level. It can be assumed that bottom-up activation of such an object can be manipulated in the same way that the perceptual salience of unimodal objects is manipulated, i.e., by increasing the relative salience (cf., Eimer et al., 2009; Lamy et al., 2004) or physical intensity of these unimodal stimuli (Theeuwes, 1994; see also Dalton & Lavie, 2004, for evidence that higher-intensity tones trigger larger auditory capture effects than lower-intensity tones). While tones of the same intensity were used in Experiment 1 and 3, the level of relative salience of the colour-change cues differed across these studies: In Experiment 1, cues appeared among uniformly grey items, whereas in Experiment 3 they were presented against a background of five differently coloured items. It is possible that tones employed across these two studies had a sufficiently high level of intensity for audiovisual synchrony to enhance the spatial bias for colour cues presented against a homogeneous, but not a heterogeneous background.

Support for this interpretation is provided by research which focused on the role of purely visual salience in the control of spatial selection of purely visual objects (e.g., Eimer & Kiss; 2008; Eimer et al., 2009; 2010; Theeuwes, 1991). Initial findings (e.g., Bacon & Egeth, 1994; Folk et al., 1992) suggested that bottom-up salience plays no role in the control of visuo-spatial attention in multi-stimulus contexts. However, more recent studies indicated that visual distractors can create a stronger spatial bias in visual selection in cases where their bottom-up activation is increased, but the effects of increased levels of visual salience on attentional orienting are typically offset by top-down inhibition (Eimer et al., 2009; Folk & Remington, 1998; Hickey et al., 2008; Lamy et al., 2003; Sawaki & Luck, 2010). The strategic suppression of processing of visual distractor events defined by nontarget features or defined on nontarget dimensions is characteristic of task contexts which require high levels of visual selectivity (Eimer & Kiss, 2008; Eimer et al., 2009; Lamy & Egeth, 2003; Lamy et al., 2004), where it is visible as a distinctive pattern of behavioural (i.e., negative cueing effects; Eimer et al., 2009; Lamy et al., 2003; 2004) and/ or electrophysiological

effects (i.e., distractor positivity; Eimer & Kiss, 2008; Hickey et al., 2008; Sawaki & Luck, 2010).

However, some studies demonstrated that highly salient task-irrelevant visual distractors can capture attention to their location even during search for objects defined by specific feature values (e.g., nontarget-colour feature singletons on fast-responses cued trials and slow-responses uncued trials, Eimer & Kiss, 2010, Experiment 2; bright and large singletons, Yantis & Egeth, 1999). For example, Yantis and Egeth (1999; Experiment 7 & 8) showed that unpredictable feature singleton distractors defined on the dimensions of brightness or size, but not colour or shape, triggered slower search times and steeper search slopes in a visual search task, in line with the notion that in some contexts salience-based capture effects cannot be offset by top-down inhibition. Extrapolating to cross-modal environments, it is possible that reliable enhancements of spatial bias by audiovisual synchrony cannot be observed unless the bottom-up activation of the emergent bimodal events exceeds a certain level. Pairing heterogeneous colour changes with the tones employed across Experiment 1 and 3 might have resulted in creation of audiovisual distractors which levels of bottom-up activation do not reach a threshold that enables irrelevant visual objects to have a reliably enhanced ability to capture attention when paired with non-visual signals.

A related explanation of the results from Experiment 3 is that multisensory integration was prevented from exerting its influence on the visual selection due to top-down inhibition of tones. The present findings are the most consistent with this account. Rather than showing no effect of tone presence, as would be expected if the salience account was correct, spatial cueing effects found on trials where sound accompanied colour cues now were numerically (but not reliably) smaller. Additionally, errors were no longer more frequent on tone-present compared to tone-absent trials. The fact that auditory events were clearly irrelevant to the task-at-hand could have motivated participants to actively suppress them.

One mechanism typically used to minimise cross-modal distraction is an increase of neural activity in the sensory cortices responsible for processing target-defining features (*intermodal attention*; Eimer & Schröger, 1998; Haxby et al., 1994; Weissman, Warner, & Woldorff, 2004; Woods, Alho, & Algazi, 1992). It is possible that participants in Experiment 3 used this form of selective attention to facilitate demanding feature-specific search in multicoloured arrays. Active suppression of neural processing in the other-modality sensory cortices, which could be responsible for the present lack of effect of tone presence on the overall RTs and spatial cueing effects, was frequently found to accompany

the strategic neural enhancement in perceptually or attentionally demanding task contexts (e.g., Hackley, Woldorff, & Hillyard, 1990; Olivers & van der Burg, 2008; Woods et al., 1992). One study provided particularly convincing evidence that in a difficult visual task, the effects of audiovisual salience in visual selection can be attenuated as a result of active inhibition of task-irrelevant tones. Olivers and van der Burg (2008) used the attentional blink paradigm (Chun & Potter, 1995; Shapiro & Raymond, 1994), in which the identification of the second target (T2) in a RSVP array is often impaired if this event is presented shortly after the first target (T1). In Olivers and van der Burg (2008; Experiment1), where the tones were paired solely with the targets in the RSVP array (i.e., sounds indicated target presence on 100% of all trials), the attentional blink was no longer observed. Contrastingly, in Experiment 3, where tones could accompany every stimulus in the array (i.e., tone validity equalled 18%), tone-induced performance facilitation was still reliable, but strongly attenuated (i.e., 10–15% vs. 3–4% in Experiment 1 and Experiment 3, respectively). Also, only now the performance for T1 was inversely correlated with performance for T2 (see Shapiro & Raymond, 1994). Literature indicates that top-down inhibition is a slow-developing (Olivers, van der Stigchel, & Hulleman, 2007) and an attentionally demanding (Watson, Humphreys, & Olivers, 2003) process. Thus, an enhanced identification of audiovisual T1 followed by impaired identification of T2 provides evidence for the cross-modal suppression account of the results from Experiment 3. .

Overall, the evidence discussed here suggests that strategic inhibition of tones, associated with high-selectivity feature-specific search mode (Bacon & Egeth, 1994) prevented the colour-change cues from capturing attention stronger when accompanied by tones in Experiment 3. Thus, a singleton-detection mode was encouraged in Experiment 4, to investigate whether decreasing the demands of the visual selection task is sufficient to enable audiovisual synchrony to boost the ability of heterogeneous colour-change cues to capture attention.

Experiment 4. The role of visual task sets in the multisensory enhancement of attentional capture in visual search

Introduction

A colour-specific task set was induced in participants in Experiment 3 in order to investigate whether multisensory integration can bias selection of objects in space through a purely bottom-up mechanism. Surprisingly, and in contrast to Experiment 1, spatial cueing effects triggered by colour cues showed now a tendency to be reduced on trials, in which the cues were accompanied by tones. These findings can be explained by activation of top-down mechanisms that facilitated performance in an attentionally demanding search task by suppressing the processing of stimuli in a task-irrelevant modality (i.e., tones), thus possibly preventing from audiovisual integration from modulating visual object selection, or preventing audiovisual integration from occurring altogether. The explanation of the present results in terms of top-down inhibition is also consistent with other cross-modal studies that involved an attentionally demanding visual selection tasks (e.g., Olivers & van der Burg, 2008).

Thus, in Experiment 4 a singleton-detection mode (Bacon & Egeth, 1994) was induced in participants to test whether, in contexts where task demands are reduced, it is possible for multisensory integration to reliably enhance spatial effects triggered by heterogeneous colour-change cues in cases where they are paired with tones. Stimuli and procedures were identical to the ones employed in Experiment 3 with a few exceptions. Similarly to Experiment 1, targets were again colour singletons (see Figure 3.3), and participants were informed that target bar colour was randomly chosen on each trial from two possible colours, e.g., blue and green. If larger spatial cueing effects were now triggered by colour-change cues on tone-present relative to tone-absent trials, this would provide evidence that the higher task demands characteristic of feature-specific search mode modulate the effects of multisensory integration on selection of visual objects in space, and that active top-down inhibition prevented the enhancement of spatial effects triggered by heterogeneous cues in Experiment 3. In contrast, if cueing effects elicited by heterogeneous colour-change cues were not modulated by tone presence even in such a low-selectivity task context, this would be consistent with the importance of the relative salience of audiovisual

synchronised distractors for their ability to create a reliably larger spatial bias than visual distractors presented alone.

Method

Participants

Twenty-two volunteers took part in this experiment (age range 19–35 years; mean age 25.3 years; 1 left-handed; 7 males). All had normal or corrected-to-normal vision and gave informed consent to participate in the study.

Stimuli, procedure, and design

The stimuli, procedures and design were identical to the ones employed in Experiment 3, with two exceptions. Namely, in order to encourage a singleton-detection mode (Bacon & Egeth, 1994), task instructions and properties of the search display were identical to the ones used in Experiment 1: Participants searched for targets of two possible colours that were kept constant for the whole study and counterbalanced between participants (8, 7, and 7 participants searched for green or blue, green or red, red or blue, respectively). Target-colour bars were presented against a uniformly grey background (see Figure 3.3). Additionally, because green was now a target-defining colour, brown (CIE chromaticity coordinates for brown: .643/.347) was used instead as one of the six task-irrelevant colours that could be assigned to the six sets of dots in the base array. Critically, one of the sets of four dots in the base array could change colour into either one of the two possible target colours (target-colour cues) or to the third colour (nontarget-colour cues) from the set of three from which the target-bar colours were chosen. For example, for participants searching for blue or green bars, changes of colour to blue or green colour represented trials with target-colour cues, and changes to red represented a nontarget-colour cue trial. In turn, for subjects searching for blue or red bars, green colour changes in the cue display represented nontarget-colour cues trials. Colour cues were paired on half of all trials with tones of the same intensity as in Experiments 1 to 3. Participants completed 12 experimental blocks with 48 trials each, resulting in a total of 576 trials.

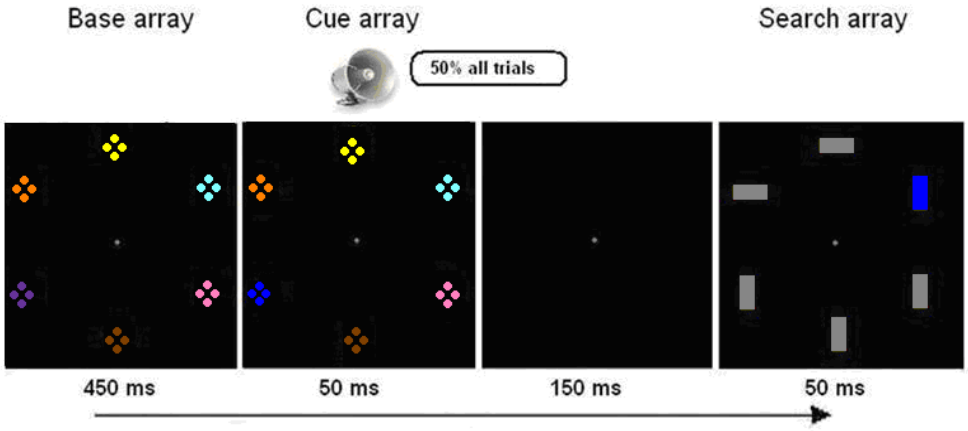

Figure 3.3. *The stimulus setup and trial sequence used in Experiment 4. Colour-change cues presented against a heterogeneous background preceded search arrays containing colour singleton target bars.*

Results

Figure 3.4 depicts average RTs for trials with correct responses and error rates, on cued and uncued trials for target-colour and nontarget-colour cues presented with and without the tones. Only data from trials on which participants responded correctly within 200 to 1000 ms from beginning of the trial were included in the analysis, which led to a loss of less than 1% of all trials. Participants missed to respond to targets also on less than 1% of all trials.

A significant main effect of spatial cueing, $F(1,21) = 21.07$, $p < .001$, $\eta_p^2 = .5$, indicated that overall participants responded faster to targets presented at cued locations relative to targets in other locations (493 ms vs. 505 ms). Importantly, this effect was not modulated by type of cue type, $F(1,21) = 1.46$, $p = .24$, indicating that cues elicited comparable cueing effects irrespective of whether they matched one of the two target colours or not. This was confirmed by follow-up tests, which revealed reliable spatial cueing effects of 14 ms for target-colour cues, $F(1,21) = 21.59$, $p < .001$, and 9 ms for nontarget-

colour cues, $F(1,21) = 5.47$, $p < .05$. There was also a main effect of tone presence, $F(1,21) = 8.33$, $p < .01$, $\eta_p^2 = .28$, evidencing that responses were faster on tone-present relative to tone-absent trials (495 ms vs. 502 ms). Critically, tone presence did not modulate the spatial cueing effects in Experiment 4 (see Figure 3.4), as suggested by a lack of a two-way tone presence by spatial cueing interaction, $F < 1$. There was also no evidence of a three-way interaction between cue type, tone presence and spatial cueing, $F(1,21) = 2.28$, $p = .15$. No other effect was significant, with only a statistical trend, $p = .09$, found for the type of cue x tone presence interaction.

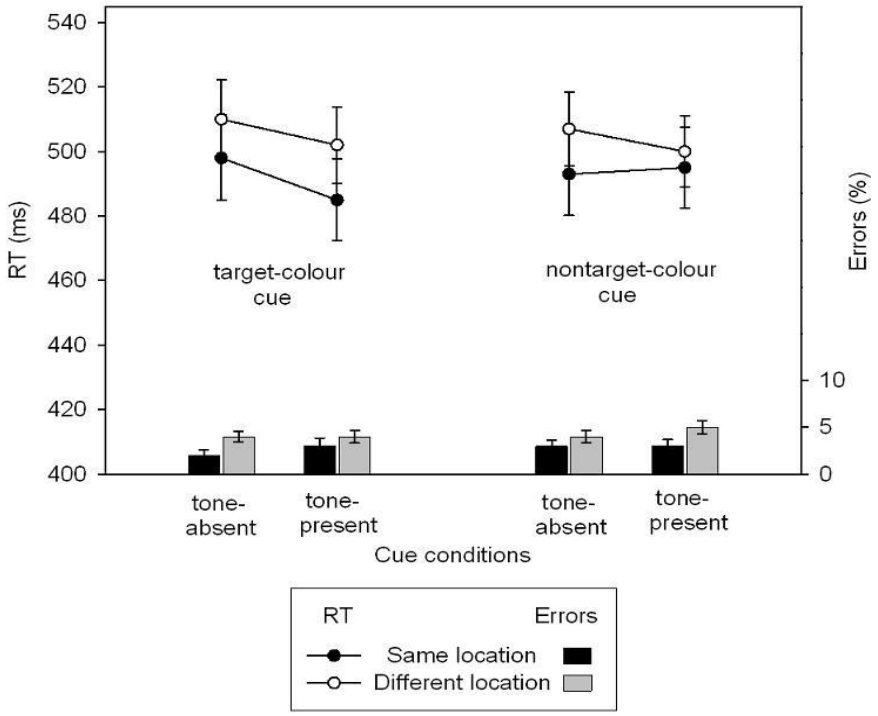

Figure 3.4. *Mean RTs (line graphs) and error rates (bar graphs) in Experiment 4 in response to targets at cued and uncued locations, shown separately for target-colour and nontarget-colour cues, and tone-present and tone-absent trials.*

Errors were less frequent on the trials on which the target was presented at a cued location when compared with trials on which targets were presented at other locations (2.8% vs. 4.2%), as suggested by a significant main effect of spatial cueing, $F(1,21) = 12.6$, $p < .01$, $\eta_p^2 = .37$. A main effect of tone presence, $F(1,21) = 5.19$, $p < .05$, $\eta_p^2 = .2$, indicated that the frequency of errors was higher on tone-present relative to tone-absent trials (3.9% versus 3.1%). No other effect was significant, with all p values > 0.2.

Across-experiment analyses

Further analyses were carried out to investigate in more detail how multisensory enhancement of attentional capture in visual search is modulated by visual selectivity and relative perceptual salience of visual cues, respectively. The first ANOVA was conducted on combined RT data from Experiments 3 and 4, for the within-subject factors of cue type, tone presence and spatial cueing, and the between-subject factor of visual selectivity (high vs. low). A lack of a three-way interaction between tone presence, spatial cueing and visual selectivity, $F < 1$, indicated that the search mode adopted by the participants did not modulate the effect of tone presence on spatial cueing effects triggered by heterogeneous colour cues . The second ANOVA was carried out on combined RT data from Experiments 1 and 4, for the within-subject factors of cue type, tone presence and spatial cueing, and the between-subject factor of cue salience (high vs. low). In contrast to visual selectivity, the level of relative salience of visual cues significantly modulated the effect of tone presence on RT cueing effects, as indicated by a three-way interaction between tone presence, spatial cueing and cue salience, $F(1,40) = 4.11$, $p < .05$, $\eta_p^2 = .09$.

Discussion

The aim of Experiment 4 was to investigate whether the level of selectivity imposed by the visual search task affects multisensory enhancement of attentional capture. For this purpose, the effect of audiovisual synchrony on spatial cueing effects triggered by colour-change cues was now tested under a singleton-detection mode (Bacon & Egeth, 1994). In spite of lower task demands in the current experiment compared to Experiment 3, the spatial effects elicited by heterogeneous colour cues were still not modulated by presence of concurrent tones. Additionally, a cross-study analysis of the combined RT data from Experiments 3 and 4 showed no evidence of visual selectivity interacting with the effect of tone presence on spatial cueing effects elicited by these cues. In other words, the ability of heterogeneous colour cues to capture attention more strongly when accompanied by task-irrelevant tones was not modulated by whether participants adopted a high-selectivity colour-specific feature search mode (in Experiment 3) or a low-selectivity search for a feature discontinuity (in Experiment 4).

The pattern of results found across Experiment 1 through 4 supports an interpretation that is consistent with the bottom-up nature of the spatial bias created by audiovisual synchrony: In Experiment 3 and 4, the level of relative salience of colour cues

presented against a heterogeneous background was not sufficiently high for multisensory integration to trigger reliably larger shifts of involuntary attention when these colour cues were paired with tones. More direct support for this particular account was provided by a cross-study analysis conducted on the RT data from Experiments 1 and 4. In contrast to visual selectivity, the relative salience of colour cues was shown to be a factor that determined whether audiovisual synchrony enhanced the ability of visual distractors to attract shifts of attention in a multi-stimulus context. Whether the colour-change cues were presented against a homogeneous background (in Experiment 1) or a heterogeneous background (in Experiment 4) was of critical importance for enhancement of spatial effects these cues triggered when accompanied by tones.

However, the difference in the effect of tone presence on spatial cueing effects found across Experiments 1 and 4 could also be accounted for by a different mechanism than the relative salience of the colour cues. In contrast to Experiment 1, homogeneous search displays were now preceded by heterogeneous cue displays, and this dissimilarity of displays could have driven the difference in results found across the two studies. Existing research (e.g., Gibson & Kelsey, 1998; Lamy et al., 2004) suggests that, similarly to a feature or dimension defining the target identity, display-wide properties of the search array in which the target is presented can also affect the ability of irrelevant objects to capture attention. Thus, it is possible that the enhancing effect of audiovisual synchrony might be reduced or even eliminated in contexts where the ability of visual cues to attract attention is already attenuated due to a lack of similarity of the distractor background against which the cues and targets appear. While possible, this explanation is inconsistent with the overall pattern of results observed across Experiments 1 to 4: If display dissimilarity was critical for the presence of multisensory enhancement of attentional capture, it should have modulated the interaction of tone presence and spatial cueing effect also when Experiment 3 and 4 were compared. In fact, no such difference was found. The results founds across these three studies are better explained by the hypothesis that the salience of bimodal distractors needs to reach a certain threshold for audiovisual synchrony to create a bottom-up bias in spatial visual selection that is reliably larger than the one triggered by purely visual distractors. This interpretation suggests that the loudness of tones used in the experiments reported so far might have been sufficiently high for higher-salience colour singleton cues (Experiment 1), but not for lower-salience heterogeneous colour cues (Experiments 3 and 4), to trigger enhanced spatial cueing effects on audiovisual trials. Thus, to directly address the bottom-up nature of the multisensory enhancement of spatial selection in vision, in Experiment 5 tones of higher intensity were paired with the heterogeneous colour-changes cues.

Experiment 5. The critical role of bottom-up salience in multisensory enhancement of visual object selection

Introduction

Experiment 1 reported in Chapter 2 demonstrated that audiovisual synchrony can enhance the ability of colour-change cues to capture attention in multi-stimulus contexts. A colour-specific feature-search mode was encouraged in Experiment 3 to test the bottom-up nature of this mechanism. However, findings from Experiment 3 and 4 indicated that bimodal distractors need to reach a certain level of perceptual salience in order for multisensory integration to enhance reliably the bias in spatial selection towards irrelevant visual objects accompanied by non-visual signals. While it is generally agreed that attentional capture by irrelevant objects is contingent on them possessing features relevant to the task-at-hand (Eimer et al., 2009, 2010; Folk et al., 1992; Lamy et al., 2004; Lien et al., 2008), evidence from visual and cross-modal research suggests that the level of relative salience of distractors can also modulate their influence on involuntary selection (Eimer et al., 2010; Olivers & van der Burg, 2008; Yantis & Egeth, 1999). Critically, the importance of the relative salience for the synchrony-driven enhancements of attentional capture was directly supported by the across-experiment analysis conducted on the data combined from Experiments 1 and 4. Considering all this evidence together, the ability of lower-salience heterogeneous colour cues should be reliably enhanced by multisensory integration in audiovisual contexts in which the relative salience of combined bimodal stimuli is enhanced.

Hence, in order to investigate the bottom-up nature of the simultaneity-induced bias in spatial selection of visual objects tones of higher intensity were paired with heterogeneous cues in Experiment 5. Aside from the intensity of the tones, all experimental procedures were identical to the ones employed in Experiment 3, i.e., observers were again encouraged to adopt a colour-specific task set. The rationale was that if the relative salience of bimodal objects is critical for audiovisual synchrony to reliably enhance the ability of visual stimuli accompanied by irrelevant tones to be selected in space via a bottom-up mechanism, then reliably larger spatial cueing effects should now be triggered by heterogeneous cues on tone-present relative to tone-absent trials. The critical question was whether this tone-induced enlargement would be observed for both target-colour and

nontarget-colour cues. If multisensory enhancement of visual attention capture is contingent on top-down settings, tone presence should only enlarge the spatial cueing effects triggered by cues matching the colour of the target. However, if multisensory integration enhances the spatial bias for visual distractors by increasing their bottom-up salience, enlarged spatial cueing effects on tone-present versus tone-absent trials should be observed for both types of colour cues.

Method

Participants

Twenty-five volunteers took part in the experiment. One participant was excluded due to their inability to perform the task as instructed, and two others because their mean RTs were more than 2 SDs longer than the group mean. The remaining 22 participants (age range 19–40 years, mean age 27.5 years; 3 right-handed; 11 males) had normal or corrected-to-normal vision. All gave informed consent to participate in the study.

Stimuli and procedure, design

Experimental procedures and analyses were identical to Experiment 3, with the sole exception that tone intensity was now higher than in Experiment 3 (80 vs. 65 dB SPL measured from a position adjacent to participant's head).

Results

Exclusion of trials with anticipatory and exceedingly slow responses led to a loss of less than 1% of all data. As shown by Figure 3.5, spatial cueing effects were now larger on trials on which the colour-change cues were accompanied by tones and this was confirmed by statistical analyses.

A main effect of spatial cueing, $F(1,21) = 47.49$, $p < .001$, $\eta_p^2 = .69$, was modulated by cue type, $F(1,21) = 39.37$, $p < .001$, $\eta_p^2 = .65$, indicating that spatial effects triggered by visual cues differed on the basis of their colour. This was confirmed by planned comparisons which showed cueing effects for target-colour cues (30 ms, $F(1,21) = 55.63$, $p < .001$, $\eta_p^2 = .73$), but a non-significant trend for nontarget-colour cues (3 ms, $F(1,21) =$

2.26, $p = .079$). Responses were faster on tone-present relative to tone-absent trials (480 ms vs. 495 ms), $F(1,21) = 32.83$, $p < .001$, $\eta_p^2 = .61$. The critical finding of Experiment 5 was that the spatial cueing effects elicited by colour-change cues were reliably enlarged on audiovisual relative to visual trials, as evidenced by a two-way interaction between tone presence and spatial cueing, $F(1,21) = 4.5$, $p < .05$, $\eta_p^2 = .18$. Pair-wise comparisons via one-tailed t-tests revealed cueing effects of 12 ms on trials on which colour-change cues were presented without a tone, $F(1,21) = 21.09$, $p < .05$, $\eta_p^2 = .18$, and cueing effects of 20 ms elicited by the same colour-change cues on trials where they were accompanied by tones, $F(1,21) = 34.09$, $p < .001$, $\eta_p^2 = .62$. As shown in Figure 3.5, the size of the enlargements of the cueing effects on audiovisual trials was very similar for both types of colour cues. Importantly, there was no interaction between spatial cueing, tone presence and cue type, $F < 1$.

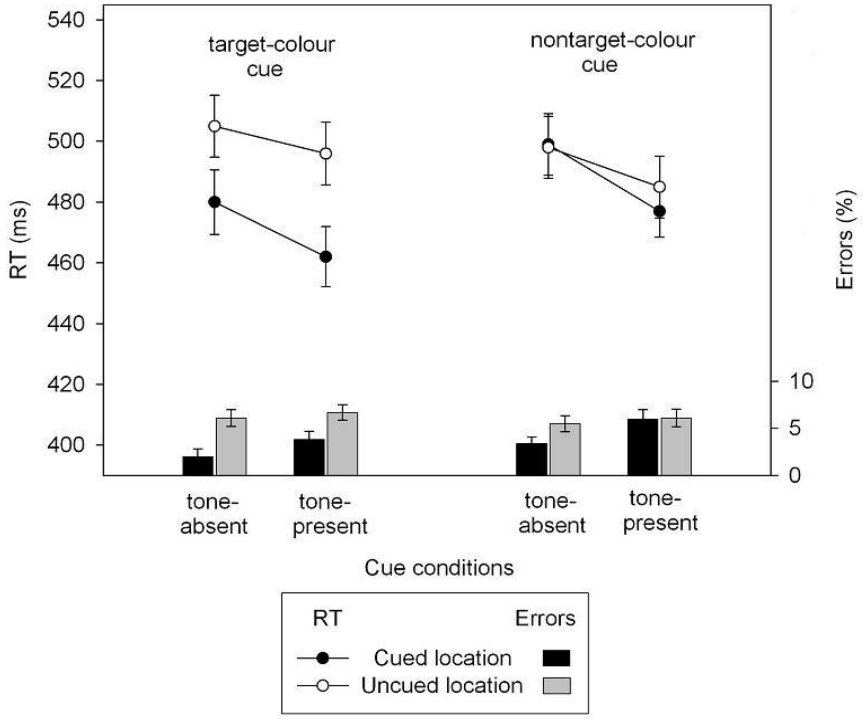

Figure 3.5. *Mean RTs (line graphs) and error rates (bar graphs) in Experiment 5 in response to targets at cued and uncued locations, shown separately for target-colour and nontarget-colour cues, and tone-present and tone-absent trials.*

Errors were more frequent on trials where target location was uncued, as opposed to trials with targets at cued locations (6.1% vs. 3.8%), evidenced by a main effect of spatial

cueing, $F(1,21) = 21.83$, $p < .001$, $\eta_p^2 = .51$. Pair-wise comparisons revealed that this difference was reliable for target-colour cues, $F(1,21) = 31.46$, $p < .001$, $\eta_p^2 = .6$, but it did not quite reach the significance level for nontarget-colour cues, $F(1,21) = 2.81$, $p = .054$, $\eta_p^2 = .12$, as reflected by a two-way interaction between cue type and spatial cueing, $F(1,21) = 21.83$, $p < .001$, $\eta_p^2 = .51$. A main effect of tone presence, $F(1,21) = 10.66$, $p < .01$, $\eta_p^2 = .34$, indicated that errors were more frequent on tone-present relative to tone-absent trials (5.7% vs. 4.3%). Tone presence also modulated the difference in error frequency between cued and uncued trials, as shown by a marginally significant two-way interaction, $F(1,21) = 4.32$, $p = .05$, $\eta_p^2 = .17$. This effect was not further modulated by cue type, $F < 1$.

Combined analysis of Experiments 3 and 5

In order to investigate in more detail the effects of tone intensity on enlargement of spatial cueing effects elicited by colour cues, one further ANOVA was conducted on the combined RT data from Experiments 3 and 5 for cue type, tone presence and spatial cueing as within-subject factors and tone intensity (low vs. high) as a between-subject factor. In line with the predictions, tone intensity was a factor that modulated the influence of tone presence on spatial cueing effects elicited by heterogeneous colour-change cues. This was demonstrated by a three-way interaction between tone intensity, tone presence and spatial cueing, $F(1,42) = 8.5$, $p < .01$, $\eta_p^2 = .17$. There was no other interaction involving tone intensity, with all p values > .18.

Discussion

Results from Experiment 5 provided the first evidence that multisensory integration can create a reliably larger bias in spatial selection of task-irrelevant visual objects accompanied by non-visual signals by increasing their bottom-up salience. RT spatial cueing effects were reliably enlarged on trials on which colour cues were accompanied by spatially diffuse tones, suggesting an enhanced ability of these cues to capture attention. In spite of participants adopting a colour-specific task set in Experiment 5, tone presence triggered larger spatial cueing effects for target-colour cues as well as for nontarget-colour cues. The fact that the multisensory enhancement of spatial object selection was not modulated by cue colour provides strong evidence that the mechanism by which audiovisual synchrony enhances spatial selection of visual objects is bottom-up in nature. Experiment 5 demonstrated that mere temporal co-occurrence of tones with visual objects can enhance the

ability of the latter to be preferentially selected in multi-stimulus contexts in circumstances where attentional capture is involuntary and rapid. Consistent with the salience-based nature of this enhancement, reliably larger RT cueing effects on tone-present relative to tone-absent trials were now observed for heterogeneous colour-change cues, what contrasts with numerically reduced cueing effects triggered by the same cues on tone-present trials in Experiment 3. This striking disparity is likely due to a higher intensity of tones employed in Experiment 5, which the cross-study analysis revealed to be of critical importance for the observed enhancements. The role of relative salience of bimodal distractors for reliably enhanced spatial effects will be discussed in more detail in the General Discussion.

Similarly to Experiment 1, faster responses and more frequent errors were found on tone-present as compared with tone-absent trials in Experiment 5. This speed-accuracy trade-off indicates that tones had an alerting influence on the performance on the visual search task. However Experiment 2 reported in Chapter 2 demonstrated that even in an experimental context designed to maximise the alerting properties of tones the observed pattern of results was more indicative of effortless integration of near-synchronous stimuli from different modalities, rather than of alertness. Additionally, the speed-accuracy trade-off was present also in Experiment 4, but there was no evidence of modulation of spatial cueing effects by tone presence in that experiment, inconsistent with a direct effect of tone-induced alerting on enhancement of attentional capture. Finally, other studies in which stimuli from different modalities were presented in close succession (e.g., Olivers & van der Burg, 2008; van der Burg et al., 2008a, 2009) also provided evidence against the importance of tone-induced alertness, by indicating that automatic integration is more common in contexts where signals from different modalities are presented in close temporal proximity. Thus, while the presence of a speed-accuracy trade-off in both experiments where multisensory enhancement of spatial effects were found suggests that a certain role for tone-induced alerting effects cannot be excluded, the overall pattern of results is consistent with a salience-based effect triggered by audiovisual synchrony.

General Discussion

The brain has developed mechanisms to bias processing of input towards those stimuli which are potentially important to our short- or long-term behavioural goals, which is often the case for objects or events that stimulate more than one sense at the same time (for a review, see Stein & Stanford, 2008). The aim of the experiments reported in Chapter 3 was

to investigate whether spatial selection can be biased towards synchronous bimodal objects on the basis of their increased perceptual salience. Surprisingly, when the effect of audiovisual synchrony on visual object selection was investigated in a context where participants adopted a colour-specific task set, heterogeneous colour-change cues triggered slightly reduced cueing effects on trials where they were accompanied by tones (Experiment 3). No simultaneity-induced enhancement was observed when spatial effects triggered by these cues were tested under a singleton-detection mode (Experiment 4). However, and most importantly, heterogeneous cues triggered larger RT spatial cueing effects on trials on which they were paired with higher-intensity tones (Experiment 5). The critical finding of Experiment 5 was that these audiovisual enlargements were not modulated by the cue colour, i.e., they were similar in size for target-colour and nontarget-colour cues.

The pattern of results observed across Experiments 3 through 5 clearly demonstrates that multisensory integration can bias attentional selection of visual objects in space by increasing the bottom-up salience of visual objects paired with task-irrelevant non-visual signals. The ability of colour cues to capture attention was enhanced when they were paired with tones, and this effect was not modulated by a top-down colour task set. In other words, spatial selection in vision can be biased towards bimodal audiovisual objects because they are more distinctive from their surroundings than unimodal visual objects. The current results are in line with a growing body of behavioural research suggesting that visual events accompanied by signals from other sensory modalities are more easily perceived (e.g., Noesselt et al., 2008; Stein et al., 1996) and oriented to more easily (e.g., Olivers & van der Burg, 2008) than when presented alone. Critically, the enhancements observed in Experiment 1 (Chapter 2) and Experiment 5 (present chapter) are consistent with numerous neurophysiological studies that indicated that temporally coincident signals from different modalities can be combined into a salient emergent multimodal event at early stages of cortical hierarchy (Fort et al., 2002a; Giard & Peronnett, 1999; Molholm et al., 2002; Murray et al., 2005). Importantly, results of tracer studies indicate that this type of multisensory convergence can occur through feedforward-type connections between primary sensory cortices (e.g., Falchier et al., 2002). The critical characteristic of these connections is that they are not modulated by feed-back projections from higher-level, heteromodal stages, which highlights impenetrability of this mechanism by endogenous attention (for a review, see Kayser & Logothetis, 2007). Consistent with this observation, enhancements of attentional capture effects by audiovisual synchrony were not modulated by top-down colour task set in Experiment 5. Further support for the bottom-up account was provided by the carried out cross-study analyses, which showed that the relative salience of emergent

audiovisual distractors, rather than the search mode adopted by participants, was of critical importance for the enhancement of the ability of these bimodal objects to attract involuntary attention when compared to purely visual distractors. Namely, when paired with lower-intensity tones, only colour-change cues presented against a homogeneous, but not heterogeneous background, triggered larger attentional capture on tone-present relative to tone-absent trials (Experiment 1 vs. Experiment 4). However, an enhanced spatial bias as function of tone presence was observed even for these lower-salience heterogeneous colour cues when higher-intensity, rather than lower-intensity, sounds were employed (Experiment 3 vs. Experiment 5). In contrast, in contexts where heterogeneous colour cues were paired with lower-intensity tones, tones did not affect visual attentional capture effects, irrespective of whether subjects adopted a high-selectivity feature-search mode or a low-selectivity singleton-detection mode (Experiment 3 vs. Experiment 4).

Overall, results from the experiments reported in Chapter 3 provide strong evidence for a direct effect of multisensory integration on spatial biases in attentional selection of visual objects in multi-stimulus contexts. However, another conclusion these findings afford is that the bottom-up spatial selection bias driven by multisensory integration differs from the one that is triggered by purely visual salience. While the perceptual salience of visual and bimodal objects might be modulated by similar factors, such as the background heterogeneity versus homogeneity of a cue array or physical intensity (e.g., lower vs. higher value on the sound pressure level scale), the enhancements found in Experiment 5 suggest that bimodal salience can affect attentional capture in an entirely bottom-up fashion. Furthermore, the results across Experiments 1 through 5 indicate that, unless the salience of the bimodal object does reach a certain threshold level, multisensory integration will have no effect on attentional object selection. Critically, once this hypothetical threshold is reached, audiovisual synchrony increases the ability of visual distractors to capture attention, irrespective of whether they share features with the target or not. The fact that this bottom-up mechanism affects spatial selection of visual objects independently of top-down attentional control settings represents its most notable property, which stands in stark contrast with the secondary, task-set contingent role of purely visual salience that has been indicated by converging electrophysiological, heamodynamic, and behavioural evidence (Downar et al., 2000; Eimer et al., 2008, 2009, 2010; Hickey et al., 2008; Lamy et al., 2003, 2004; Sawaki & Luck, 2010; Serences et al. 2001).

While the results reported in Chapter 3 provide evidence for a novel mechanism of bottom-up control of visuo-spatial attention in multi-stimulus contexts, they also indicate that its relative role might be small when compared with the effects of local feature contrasts

within a specific modality. In the experiments, where audiovisual enhancements of attentional capture were found (Experiments 1 and 5), these enhancements were considerably smaller than the spatial cueing effects that were triggered by colour cues presented alone (10 ms vs. 25 ms). Additionally, these cross-modal effects were similar in size across the two experiments, in spite of the fact that combinations of stimuli of quite different levels of perceptual salience were employed. In the light of the visual salience results, the range of these audiovisual enhancements seems rather limited. Thus, it would be important for the generalisability of the current findings to investigate whether larger multisensory enhancements of attentional capture could be observed in multi-stimulus contexts, e.g., in an experimental context in which the time interval between the cues (visual and audiovisual) and target is not predictable for observers. This possibility is suggested by a comparison of two studies that both investigated the role of visual salience in the ability of target-matching distractors to capture attention, but differed with respect to the predictability of the cue-target time interval. In a study where cues preceded targets by four randomly intermixed stimulus-onset asynchronies (Lamy et al., 2004; Experiment 1), numerically larger RTs spatial cueing effects were triggered by both singleton (55 ms vs. 39 ms) and heterogeneous (51 ms vs. 31 ms) target-colour cues when compared with a study (Eimer et al., 2009; Experiment 1) where a single stimulus-onset asynchrony was used. This difference suggests an important role of temporal predictiveness of distractors for the salience-based effects in attentional capture. It is therefore possible that larger audiovisual enhancements might have been observed if unpredictable cue-time intervals had been used in Experiment 1 (Chapter 2) and Experiment 5 (Chapter 3) reported in the current thesis.

In summary, the evidence presented in Chapter 3 demonstrates that multisensory integration can have a direct effect on attentional object selection in multi-stimulus environments by increasing perceptual salience of visual objects accompanied by non-visual signals. To provide a better understanding of how visual attention capture effects based on audiovisual salience differ from those based on purely visual salience, the last two experiments of this part of the thesis, described in Chapter 4, investigated how audiovisual synchrony modulates neural correlates of visual object selection.

Chapter 4. Electrophysiological evidence for multisensory enhancement of a bottom-up bias in visual object selection

The synchrony-based enhancements of spatial selection of irrelevant visual objects accompanied by irrelevant tones that were observed in Experiment 5 (Chapter 3) are consistent with the existing neurophysiological and behavioural research that has provided initial evidence for a direct effect of multisensory integration on the competition that occurs naturally among visual stimuli in multi-stimulus context. Multisensory integration was shown to increase the ability of visual objects to capture attention more strongly when accompanied by non-visual signals (Olivers & van der Burg, 2008; Vroomen & de Gelder, 2000; see Koelewijn et al., 2010, for a review). However, mixed results (i.e., larger attentional capture by audiovisual than visual distractors in the majority, but not all, SOA conditions in van der Burg et al. [2008a, Experiment 5]) or methodological issues (i.e., a lack of a visual-distractor condition against which the magnitude of attentional capture by audiovisual distractors could be compared in van der Burg et al. [2008b; Experiment 3]), prevented the previous studies from providing convincing evidence for a bottom-up selection bias towards visual objects paired with non-visual signals triggered by multisensory integration. Thus, Experiment 5 reported in Chapter 3 is the first behavioural study that directly supported the idea that in multi-stimulus settings audiovisual synchrony can enhance visual attention capture via a salience-based mechanism (see Matusz & Eimer, 2011).

Nevertheless, the nature of the evidence provided by Experiment 5 in respect to the bottom-up character of multisensory enhancement of visual object selection is still somewhat indirect: The mechanism in question was investigated by measuring behaviourally how the speed of visual target selection is modulated by the location of visual versus audiovisual cues preceding that target on every trial. While the hypothesis that tone-induced alertness is the critical mechanism responsible for the observed enhancement of RT spatial cueing effects was falsified in Experiment 2 (Chapter 2), a certain role of transient alertness in enlargements of behavioural cueing effects cannot be ruled out (cf., van der Burg et al., 2008a, for a similar argument with respect to the 'pip-and-pop' effect). Thus, a more direct demonstration of modulations of visual selection by audiovisual synchrony

would further strengthen the interpretation of this effect in terms of a synchrony-induced increase of bottom-up salience of visual objects paired with non-visual signals. The 'N2-posterior-contralateral' or 'N2pc' component, regarded traditionally as an online marker of attentional selection (Eimer, 1996; Kiss & Eimer, 2011; Luck & Hillyard, 1994a), has proved to be particularly useful in investigating the role of purely visual salience in attentional selection (e.g., Eimer et al., 2009, 2010). Because the N2pc component can be triggered by targets as well as potential target stimuli (i.e., distractors that possess task-relevant features and distractors that are highly distinctive from their background; for more details see Section 1.4.3.3), this visual ERP component is well suited for investigations of the bottom-up nature of the visual selection bias triggered towards synchronous audiovisual objects by multisensory integration.

An important related insight that can be provided by the N2pc component into synchrony-induced enhancement of a bottom-up bias in spatial selection is the exact neural mechanism by which salient irrelevant audiovisual objects are preferentially selected in vision: The neural mechanism driving the effects of audiovisual salience in visual object selection may be visible as onset latency effects, amplitude effects or a combination of both. The effects of audiovisual salience might involve speeding up of attentional orienting (e.g., Bell et al., 2005, Wallace, Meredith, & Stein, 1998), where competition in vision may be biased towards synchronised audiovisual objects because their perceptual processing is completed faster compared to purely visual objects. Alternatively, audiovisual salience may primarily affect N2pc amplitudes (van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011), having triggered enhanced neural responses to audiovisual versus visual objects in sensory-perceptual visual cortices during feedforward processing (cf., Lakatos et al., 2005, 2007). Pinpointing the exact neural marker associated with the effects of audiovisual salience in visual object selection will further our understanding of the possible differences between neuro-cognitive mechanisms responsible for cortex-mediated salience-based effects of multimodal synchrony and SC-mediated effects that underlie the multisensory facilitation of orienting to faint stimuli in the periphery (cf., Stein, 1998).

Thus, the aim of the experiments reported in Chapter 4 was to provide direct evidence for a bottom-up bias in visual attentional selection towards salient audiovisual objects by measuring whether and how specifically the N2pc component triggered in response to task-irrelevant visual distractors is modulated by concurrent presentation of spatially diffuse tones. For this purpose, participants in Experiment 6 were again searching for target bars of one specific colour (e.g., 'red bars'; see Experiment 5 in Chapter 3), while behavioural (RT spatial cueing effects) and electrophysiological (the N2pc component)

indices of attentional capture were measured in response to heterogeneous as well as singleton colour-change cues. Surprisingly, mixed results were obtained in this experiment, with no evidence for synchrony-driven enlargements of behavioural spatial cueing effects, and synchrony-driven enhancements of cue-elicited N2pc amplitudes observed only for singleton colour cues. To replicate and extend the modulations of the N2pc component by audiovisual salience, in Experiment 7 participants were searching for targets defined as a conjunction of specific visual and auditory features (e.g., a blue bar accompanied by a high-pitch tone). Such experimental setting provided means to assess whether salience-based multisensory enhancements of visual objects selection can be observed also in contexts where attentional control is set also for auditory features. Additionally, to eliminate the complexities associated with the sequential presentation of the cue and search arrays, in Experiment 7 no cue arrays were shown and the auditory stimuli were presented synchronously with the visual stimuli in the search array.

Experiment 6. The N2pc component as the ERP marker of the salience-based audiovisual bias in visual object selection

Introduction

In the visual domain, the N2pc component proved to be extremely useful to study the relative hierarchy between top-down and bottom-up factors in the control of visuo-spatial attention in multi-stimulus contexts. Several experiments (e.g., Eimer et al., 2009, 2010, Kiss et al., 2012) have helped to reconcile the contradictory results that purely behavioural studies (Folk et al., 1992, 1998; Theeuwes, 1991, 1994; Theeuwes et al., 2000) provided in respect to the automaticity of attentional capture by salient but task-irrelevant visual objects. According to a recent two-stage selection model proposed by Kiss and colleagues (2013), behavioural spatial cueing effects are indicative of whether attentional focus is maintained at the location of the cue until the moment of target stimulus presentation. In contrast, the N2pc component reflects the initial stage of attentional selection, where the activity on a hypothetical salience map (Wolfe, 1994, 2007) is computed for each location in external space. Thus, studies that have used a combination of behavioural and electrophysiological measures of attentional capture enriched our understanding of the relative role of purely visual salience in the control of attentional selection by providing an insight into the effects of visual salience that may not necessarily result in a bias sufficiently strong to be reflected by behavioural cueing effects.

In one such study, Eimer et al. (2009) investigated how the level of relative salience affects the ability of target-matching colour cues to capture attention. Behavioural spatial cueing effects triggered by target-colour singleton cues and heterogeneous colour cues (cf., Experiment 3 in Chapter 3) were comparable in size (i.e., there was only a trend for larger cueing effects for singleton colour cues; see also Lamy et al., 2004). However, when the N2pc components elicited by these two types of target-colour cues were compared, mean N2pc amplitudes triggered in response to the higher-salience singleton colour cues were reliably enhanced compared to the lower-salience heterogeneous colour cues, suggesting that higher levels of salience may not always result in a reliably stronger behavioural capture effects. In another study, Eimer and Kiss (2010; Experiment 2) showed that in contexts where stringent colour-specific top-down task set prevented nontarget colour singleton cues

from eliciting reliable RTs spatial cueing effects, a weak but reliable N2pc component was still observed in response to these cues. This result likely highlights fluctuations in the maintenance of attentional control that might have not been strong enough for the effects of visual salience on attentional selection to be reliably observed with behavioural measures.

The studies that combined ERP and behavioural indices of attentional capture not only provided a more detailed picture on the relative importance of visual salience in the control of attentional capture by task-irrelevant distractors, but they also revealed that N2pc amplitude effects are primarily associated with increased levels of visual salience (e.g., Eimer et al., 2009, 2010). Notably, the findings from these studies and the behavioural results from Experiment 5 (Chapter 3) demonstrate a striking contrast between the bottom-up biases in visual competition created by visual and audiovisual salience, as the latter can operate in a fashion entirely independent of the current top-down control settings. This contrast poses a question whether it is paralleled by a difference in the neural mechanisms underlying these two forms of bottom-up bias in visual selection: While only N2pc amplitude effects are typically found for salient but task-irrelevant visual distractors, for salience effects driven by audiovisual synchrony a pattern of N2pc results defined by, for example, both onset latency and amplitude effects may be characteristic.

The existing multisensory literature does not provide sufficient evidence for the type of neural mechanism by which multisensory integration can enhance the ability of task-irrelevant visual distractors to capture attention in multi-stimulus contexts. On the one hand, the salience effects triggered by bimodal synchrony may be typically reflected in 'redundancy gains' that speed up the process in question, irrespective of whether this is attentional orienting (e.g., Bell et al., 2005, Wallace, Meredith, & Stein, 1998) or a manual response (e.g., Gondan, Goetze, & Greenlee, 2010; Gondan, Niederhaus, Rösler, & Röder, 2005). In multi-sensory contexts, visual competition may be biased towards synchronised audiovisual objects because their perceptual processing is completed faster compared to purely visual events. On the other hand, multisensory integration may primarily affect N2pc amplitudes. This account is supported by neurophysiological studies which have demonstrated that the other-modality signal resets the phase of the ongoing oscillatory activity in the primary-modality sensory cortices, resulting in the processing of the primary stimulus at the point of maximal activity (Ghazanfar & Chandrasekaran, 2007; Lakatos et al., 2005, 2007).

Some initial evidence for N2pc amplitude enhancement as the neural mechanism by which audiovisual salience modulates visual object selection was provided by the only study that investigated whether the N2pc component can be modulated by audiovisual synchrony.

In a replication of the 'pip-and-pop' phenomenon (cf., van der Burg et al., 2008a, Experiment 5), van der Burg and colleagues (2011) compared ERPs elicited by lateralised visual distractors presented on tone-present versus tone-absent trials. A reliable N2pc component was triggered in response to audiovisual distractors, but not by the same distractors presented without a tone. These results are in line with multisensory integration providing task-irrelevant visual objects accompanied by non-visual signals with a competitive advantage that is based on increased bottom-up salience. However, the N2pc component observed in the tone-present condition was reliable only during a time window of 20 ms in length (between 210 and 230 ms after distractor onset), which is unusually short when compared to N2pc components typically found in the literature (e.g., Astle, Nobre, & Scerif, 2010; Eimer, Kiss, & Nicholas, 2011; Luck & Hillyard, 1994; Sawaki & Luck, 2010). Critically, no evidence was provided that mean N2pc amplitudes triggered in response to visual versus bimodal distractors were reliably different, what precludes treating these findings as strong evidence in support of modulations of the N2pc components by audiovisual salience, leaving open also the issue of the neural mechanism by which this modulation takes place.

To provide direct evidence for audiovisual synchrony as a mechanism of bottom-up bias in visual selection and discover the neural marker of this selection bias, Experiment 6 used methods similar to Experiment 5 (Chapter 3), but now attentional capture by visual versus audiovisual cues was measured with both behavioural and ERP methods. If reliable audiovisual modulations of the N2pc component accompany corresponding enhancements of behavioural capture, this would be a clear demonstration that multisensory integration can be a source of bottom-up bias in attentional selection in vision. Also, blocks with singleton colour cues were included in Experiment 6, to investigate whether stronger effects of audiovisual synchrony on both behavioural and ERP measures are triggered by visual distractors of higher relative salience (cf., General Discussion in Chapter 3). Thus, half of all experimental blocks were identical to the ones employed in Experiment 5 (Chapter 3), i.e., heterogeneous search displays were preceded by heterogeneous cue displays. The experimental procedures in the remaining blocks were identical, with the sole exception that now colour singletons were always presented in the cue array. The critical prediction in Experiment 6 was that audiovisual enhancements of visual attention capture should be visible not only as enlargements of behavioural spatial cueing effects (cf., Experiment 5 in Chapter 3), but also as enhancements of the cue-induced N2pc components (van der Burg et al., 2011), when colour cues are compared on tone-present and tone-absent trials. Another prediction was that stronger enhancements of attentional capture, as measured by the N2pc

component (cf., Eimer et al., 2009) as well as, potentially, RT behavioural spatial cueing effects, will be observed in blocks where cues are colour singletons compared to heterogeneous colour cues are paired with irrelevant tones, as higher levels of visual salience should result in an overall larger spatial bias in visual selection following audiovisual integration.

Method

Participants

Eighteen volunteers took part in this experiment. Two were excluded due to excessive eye movements. All the remaining ones (age range 19–31 years; mean age 24.6 years; 1 left-handed; 7 males) had normal or corrected-to-normal vision, gave informed consent to participate in the study.

Stimuli, procedure, and design

Similarly to Experiment 5 from Chapter 3, participants searched for target bars of specific pre-defined colour among multiple coloured distractor bars. On each trial, the search array preceded by a cue array containing a change to a target or nontarget colour which could be accompanied by a spatially diffuse tone (see Figure 3.1 in Chapter 3). In the blocks with heterogeneous colour cues, the stimuli and experimental procedures were identical to the ones employed in Experiment 5, except that the number of blocks for each hand-key mapping (e.g., 'top button - left index finger, bottom button - right index finger') was now increased from four to five, resulting in a total of 640 experimental trials across 10 blocks. Similar procedures were used in the blocks with singleton colour cues, with the exception that now the cue was a unique colour change that was presented against a uniformly grey background. The colour of the target bar and the order of blocks were fully counterbalanced across participants.

EEG recording and data analysis

EEG was DC-recorded from 23 scalp electrodes mounted in an elastic cap at standard positions of the extended 10-20 system at sites, Fpz, Fz, F3, F4, F7, F8, FC5, FC6, Cz, C3, C4, T7, T8, CP5, CP6, Pz, P3, P4, P7, P8, PO7, PO8 and Oz (500Hz sampling rate; 40Hz

low-pass Butterworth filter). All scalp electrodes were online referenced to the left earlobe and re-referenced offline to the average of both earlobes. Impedances were kept below 5 kΩ. Horizontal eye movements (HEOG) were measured from two electrodes placed at the outer canthi of the eyes. Only trials with correct responses to targets were analyzed. Trials with saccades (voltage exceeding ± 30 µV in the HEOG channel), eyeblinks (exceeding ± 60 µV at Fpz) or muscle artefacts (exceeding ± 80 µV at any other electrode) were excluded from the analyses, as were trials with incorrect responses and missed targets.

EEG in response to cue stimuli was epoched and averaged for the 500 ms interval after cue onset, relative to a 100 ms pre-cue baseline. Averages were computed for trials with colour cues in the left and right hemifield, separately for target-colour and nontarget-colour cues, and for tone-present and tone-absent trials. Mean N2pc amplitudes quantified on the basis of ERP activity elicited between 170 ms and 270 ms after cue onset at lateral posterior electrodes PO7 and PO8 were analysed in a repeated-measures ANOVA for the factors cue type (target-colour versus nontarget-colour cue), tone presence (tone present versus tone absent) and contralaterality (electrode ipsilateral versus contralateral to the colour cue). Analyses were conducted separately for blocks with heterogeneous and singleton colour cues.

Results

Behavioural performance

Blocks with heterogeneous colour cues

Participants gave anticipatory or extremely slow responses on less than 1.5% of all trials, and failed to respond on less than 1% of trials. Similarly to Experiment 5 from Chapter 3, RT spatial cueing effects triggered by heterogeneous colour cues differed as a function of cue colour (see Figure 4.1). This was evidenced by a main effect of spatial cueing, $F(1,15) = 50.13$, $p < .001$, $\eta_p^2 = .77$, modulated by cue type, $F(1,15) = 8.82$, $p < .01$, $\eta_p^2 = .37$. Planned comparisons demonstrated that cueing effects elicited by nontarget-colour cues were strongly attenuated albeit still significant (8 ms, $F(1,15) = 6.26$, $p < .05$, $\eta_p^2 = .29$) when compared to target-colour cues (27 ms, $F(1,15) = 36.15$, $p < .001$, $\eta_p^2 = .71$). Responses were faster on tone-present relative to tone-absent trials (513 ms vs. 527 ms, $F(1,15) = 22.24$, $p < .001$, $\eta_p^2 = .59$. In line with our predictions, RT cueing effects triggered by heterogeneous colour cues were numerically larger on tone-present relative to tone-absent

trials (21 ms vs. 14 ms; see Figure 4.1). However, this enlargement was not reliable, with no evidence of a tone presence by spatial cueing interaction, $F(1,15) = .96$, $p = .17$. Importantly, the absence of this two-way interaction was not driven by presence of a three-way interaction involving cue type, $F(1,15) = 1.2$, $p = .28$.

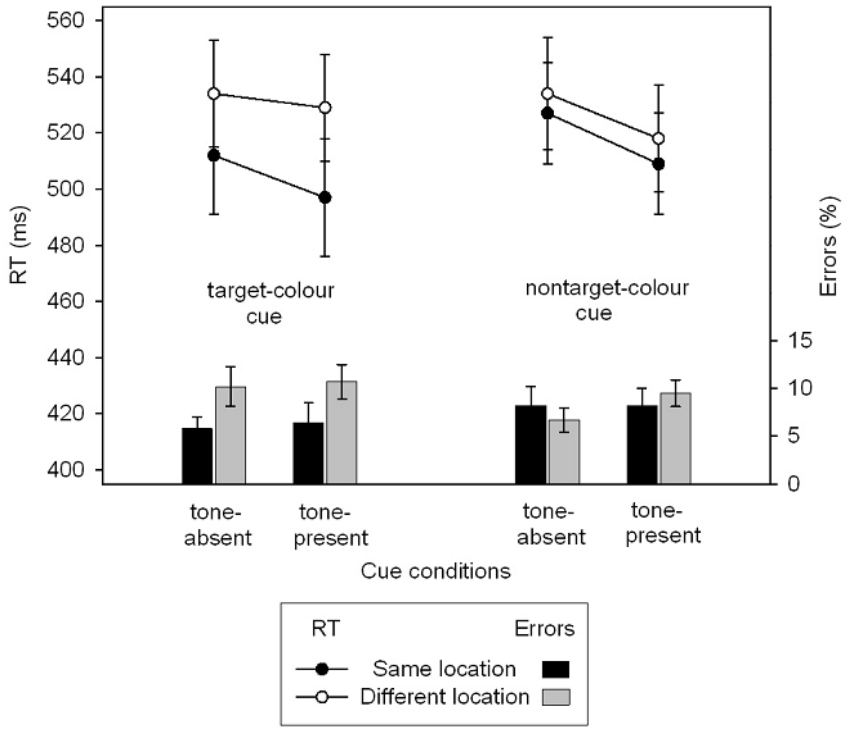

Figure 4.1. *Mean RTs (line graphs) and error rates (bar graphs) observed in Experiment 6 in the heterogeneous-cue blocks in response to targets at cued and uncued locations, shown separately for target-colour and nontarget-colour cues, on tone-present and tone-absent trials.*

Analysis of error rates showed a main effect of spatial cueing, $F(1,15) = 9.44$, $p < .01$, $\eta_p^2 = .39$, modulated by cue type, $F(1,15) = 10.06$, $p < .01$, $\eta_p^2 = .4$. Pair-wise comparisons confirmed the presence of reliable differences between error rates on cued and uncued trials for target-colour cues (6.2% vs. 10.4%), $F(1,15) = 14.83$, $p < .001$, $\eta_p^2 = .5$, but not for nontarget-colour cues, $F < 1$. Cueing effects visible on error rates were not modulated by tone presence, $F(1,15) = 1.58$, $p = .23$, and there was also no evidence of a three-way interaction involving cue type, $F < 1$.

Blocks with singleton colour cues

The overall proportion of trials with missed, premature or exceedingly slow responses in the blocks with singleton colour cues was smaller than 1%. As visible in Figure 4.2, spatial cueing effects again differed on the basis of cue colour, what was evidenced by a main effect of spatial cueing, $F(1,15) = 25.98$, $p < .001$, $\eta_p^2 = .63$, modulated by cue type, $F(1,15) = 27.11$, $p < .001$, $\eta_p^2 = .64$. Pair-wise comparisons confirmed this by showing that only singletons that matched the target colour elicited reliable cueing effects (30 ms, $F(1,15) = 53.6$, $p < .001$, $\eta_p^2 = .78$), while singletons of a nontarget colour failed to do so ($F < 1$). Responses were faster on trials on which target bars were preceded by target-colour cues compared to nontarget-colour cues (498 ms vs. 507 ms), $F(1,15) = 15.4$, $p < .001$, $\eta_p^2 = .51$, and on tone-present compared to tone-absent trials (499 ms vs. 506 ms), $F(1,15) = 11.8$, $p < .01$, $\eta_p^2 = .44$. More importantly, even in blocks where cues were colour singletons spatial cueing effects were not enlarged on trials on which the colour cues were accompanied by tones, as shown by a lack of two-way interaction between tone presence and spatial cueing, $F < 1$. Again, there was also no three-way interaction involving cue type, $F < 1$.

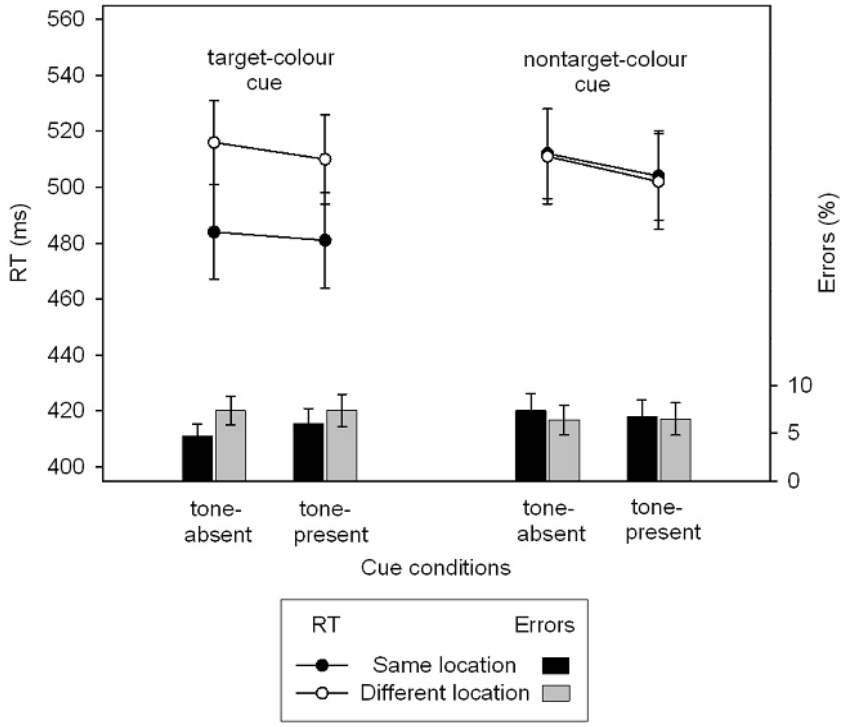

Figure 4.2. *Mean RTs (line graphs) and error rates (bar graphs) observed in Experiment 6 in the singleton-cue blocks in response to targets at cued and uncued locations, shown separately for target-colour and nontarget-colour cues, on tone-present and tone-absent trials.*

Analysis of error rates again revealed a main effect of spatial cueing, $F(1,15) = 3.1$, $p < .05$, $\eta_p^2 = .16$, modulated by cue type, $F(1,15) = 6.61$, $p < .05$, $\eta_p^2 = .29$, indicating that also in singleton-cue blocks spatial cueing effects on error rates were determined by cue colour. Planned comparisons confirmed this by demonstrating fewer errors on cued relative to uncued trials, but only for target-colour cues (5% vs.7.6%, $F(1,15) = 7.87$, $p < .05$, $\eta_p^2 = .33$). There was no such difference for nontarget-colour cues ($F < 1$). Cueing effects measured on error rates were not modulated by tone presence or an interaction between cue type and tone presence, both F's < 1.

Across-block comparison

To provide a better understanding of the pattern of divergent N2pc results found across heterogeneous-cue and singleton-cue blocks (described below), a four-way repeated-measures ANOVA for the factors cue type, tone presence, spatial cueing and cue salience (low vs. high) was conducted separately on RTs and error rates. The RTs data showed a non-significant tendency for overall slower responses in heterogeneous-cue relative to singleton-cue blocks, $F(1,15) = 3.26$, $p = .09$. The RT difference between tone-present versus tone-absent trials was reliably larger in blocks with heterogeneous colour cues (13 ms) than for singleton cues (7 ms), as indicated by a two-way tone presence x relative salience interaction, $F(1,15) = 5.27$, $p < .05$, $\eta_p^2 = .26$. The modulation of this effect by spatial cueing just missed the statistical significance threshold, $F(1,15) = 1.14$, $p = .054$, $\eta_p^2 = .23$, suggesting that RTs cueing effects elicited by heterogeneous colour cues tended to be more enhanced by tone presence than the corresponding effects for singleton colour cues. The error rates analysis showed more frequent errors in blocks with heterogeneous colour cues compared to singleton cues, $F(1,15) = 8.37$, $p < .05$, $\eta_p^2 = .36$. There was also a trend for a two-way spatial cueing x cue salience interaction, $F(1,15) = 3.89$, $p = .067$, $\eta_p^2 = .21$, indicating reliable error-rates cueing effects for heterogeneous colour cues (2.1%, $p < .01$), but not for singleton colour cues, $F < 1$. There was no other interaction on RTs or error rates involving cue salience as a factor.

N2pc results

Blocks with heterogeneous colour cues

Figure 4.3 (top panel) shows ERPs triggered in response to heterogeneous cue arrays in the 350 ms interval after cue onset at electrodes PO7/8 contralateral and ipsilateral to the side of target-colour and nontarget-colour cues, separately for tone-absent and tone-present trials. Analysis of the N2pc mean amplitudes showed no significant main effect of contralaterality, $F < 1$. There was only a weak trend for a two-way cue type x contralaterality interaction, $F(1,15) = 1.43$, $p = .12$, suggesting a tendency for cue-elicited N2pc elicited by nontarget-colour cues to be attenuated compared to target-colour cues. A main effect of tone presence, $F(1,15) = 7.96$, $p < .05$, $\eta_p^2 = .35$, reflected a reduction of N2 amplitudes on trials where colour cues were accompanied by tones. As visible in Figure 4.3 (bottom panel), the difference waveforms, which were obtained by subtracting the ipsilateral from contralateral ERPs, were smaller on tone-present relative to tone-absent trials, but this reduction was not reliable, as demonstrated by a lack of two-way tone presence by contralaterality interaction, $F(1,15) = 1.6$, $p = .22$. Importantly, this null effect was not driven by a three-way interaction involving cue type, $F(1,15) = 2.23$, $p = .16$.

Blocks with singleton colour cues

Figure 4.4 (top panel) depicts ERPs triggered in response to singleton cue arrays in the 350 ms interval after cue onset at electrodes PO7/8 contralateral and ipsilateral to the side of target-colour and nontarget-colour cues, separately for tone-absent and tone-present trials. In contrast to heterogeneous cues, the N2pc components elicited by singleton colour cues were rendered to be larger for audiovisual relative to visual cues. This is clearly visible in the difference waveforms shown in Figure 4.4 (bottom panel).

Analysis of the N2pc mean amplitudes in these blocks revealed that the N2pc components triggered by singleton colour cues were modulated by cue colour, as indicated by a main effect of contralaterality, $F(1,15) = 14.85$, $p < .01$, $\eta_p^2 = .5$, that was accompanied by a two-way cue type by contralaterality interaction, $F(1,15) = 7.65$, $p < .05$, $\eta_p^2 = .34$. Planned comparisons showed that an attenuated albeit reliable N2pc was triggered by nontarget-colour cues (-.05 µV, $F(1,15) = 9$. $p < .01$, $\eta_p^2 = .38$) as well as by target-colour cues (-1.1 µV, $F(1,15) = 15.64$, $p < .001$, $\eta_p^2 = .51$).

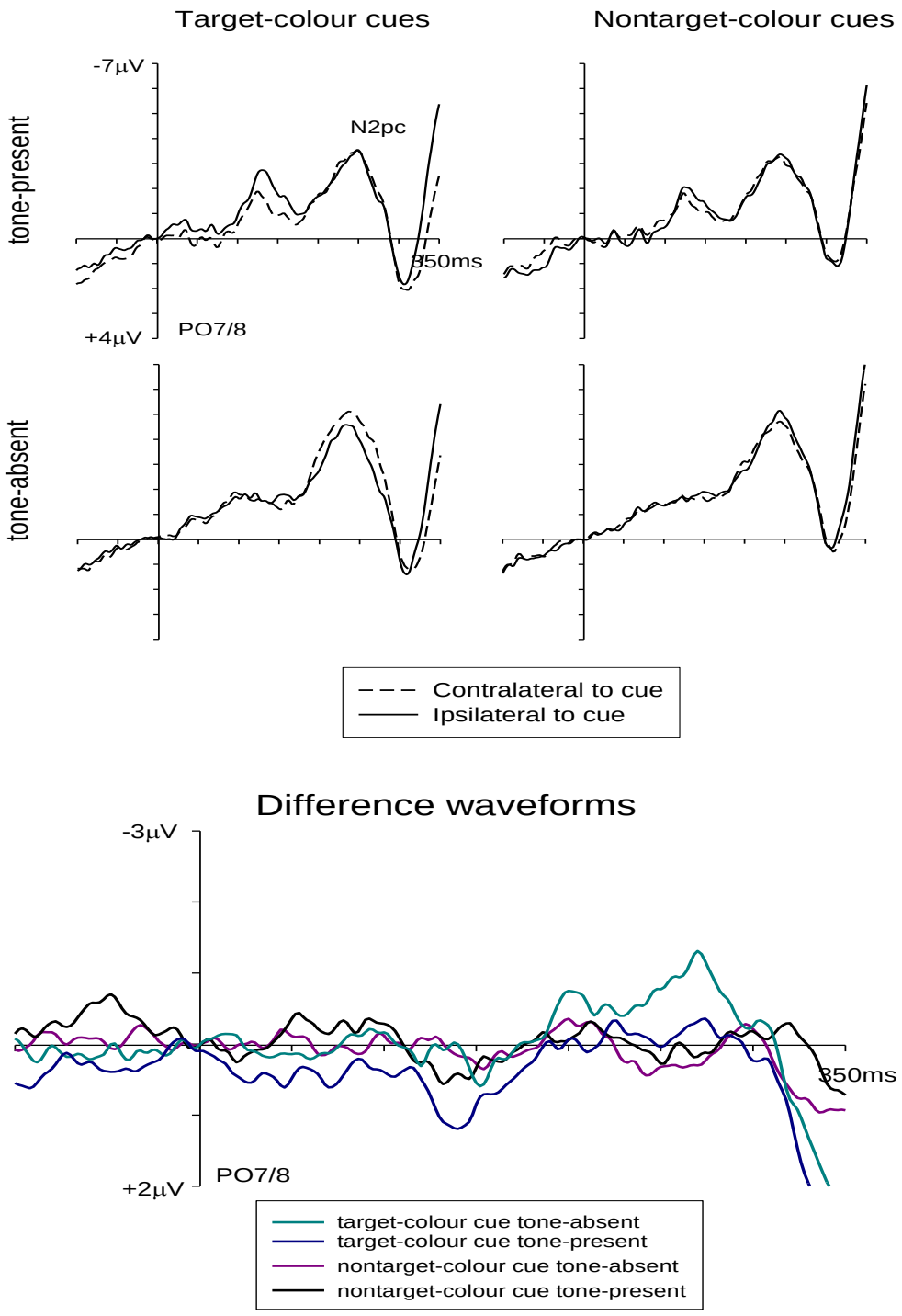

Figure 4.3. *Top panel: Grand averaged visual ERPs obtained in Experiment 6 at electrodes PO7/PO8 contralateral and ipsilateral in response to heterogeneous colour cue arrays, shown separately for target-colour and nontarget-colour cues and for tone-present and tone-absent trials. Bottom panel: Difference waveforms computed by subtracting ipsilateral from contralateral ERPs elicited by heterogeneous colour cues, shown separately for target-colour and nontarget-colour cues, and tone-present and tone-absent trials.*

A main effect of tone presence was also found, $F(1,15) = 16.05$, $p < .001$, $\eta_p^2 = .52$, reflecting smaller N2 components on tone-present relative to tone-absent trials in the singleton-cue blocks. Critically and in line with our predictions, the N2pc components triggered in response to singleton colour cues were enhanced on trials on which these cues were presented concurrently with tones, evidenced by a two-way interaction between tone presence and contralaterality, $F(1,15) = 3.23$, $p < .05$, $\eta_p^2 = .18$. These enhancements did not differ between target- and nontarget-colour cues, $F < 1$.

Across-block comparison

To investigate whether the differences in salience of colour cues were responsible for the divergent pattern of ERP results observed for blocks with heterogeneous and singleton cue arrays, a four-way repeated-measures ANOVA for the factors cue type, tone presence, spatial cueing and cue salience (low vs. high) was carried out on the N2pc mean amplitudes measured across the two types of experimental blocks. The results revealed that the N2pc components were overall larger for singleton cues as compared to heterogeneous colour, evidenced by a two-way cue salience x contralaterality interaction, $F(1,17) = 11.04$, $p < .01$, $\eta_p^2 = .39$. Critically, this interaction was further modulated by tone presence, $F(1,17) = 6.34$, $p < .05$, $\eta_p^2 = .27$.

Two further ANOVAs were conducted to directly compare the N2pc components elicited by heterogeneous and singleton colour cues on purely visual and audiovisual trials, respectively. On tone-absent trials, there was no difference in N2pc amplitudes elicited by singleton and heterogeneous cues, $F(1,17) = 2.53$, $p = .13$. In contrast, on tone-present trials singleton colour cues elicited enhanced N2pc components compared to heterogeneous colour cues, $F(1,17) = 17.9$, $p < .001$, $\eta_p^2 = .51$. This result was due to the fact that a reliable N2pc component was triggered in response to audiovisual singleton colour cues, $F(1,17) = 11.4$, $p < .01$, $\eta_p^2 = .4$, but no N2pc was found for audiovisual heterogeneous colour cues, $F < 1$.

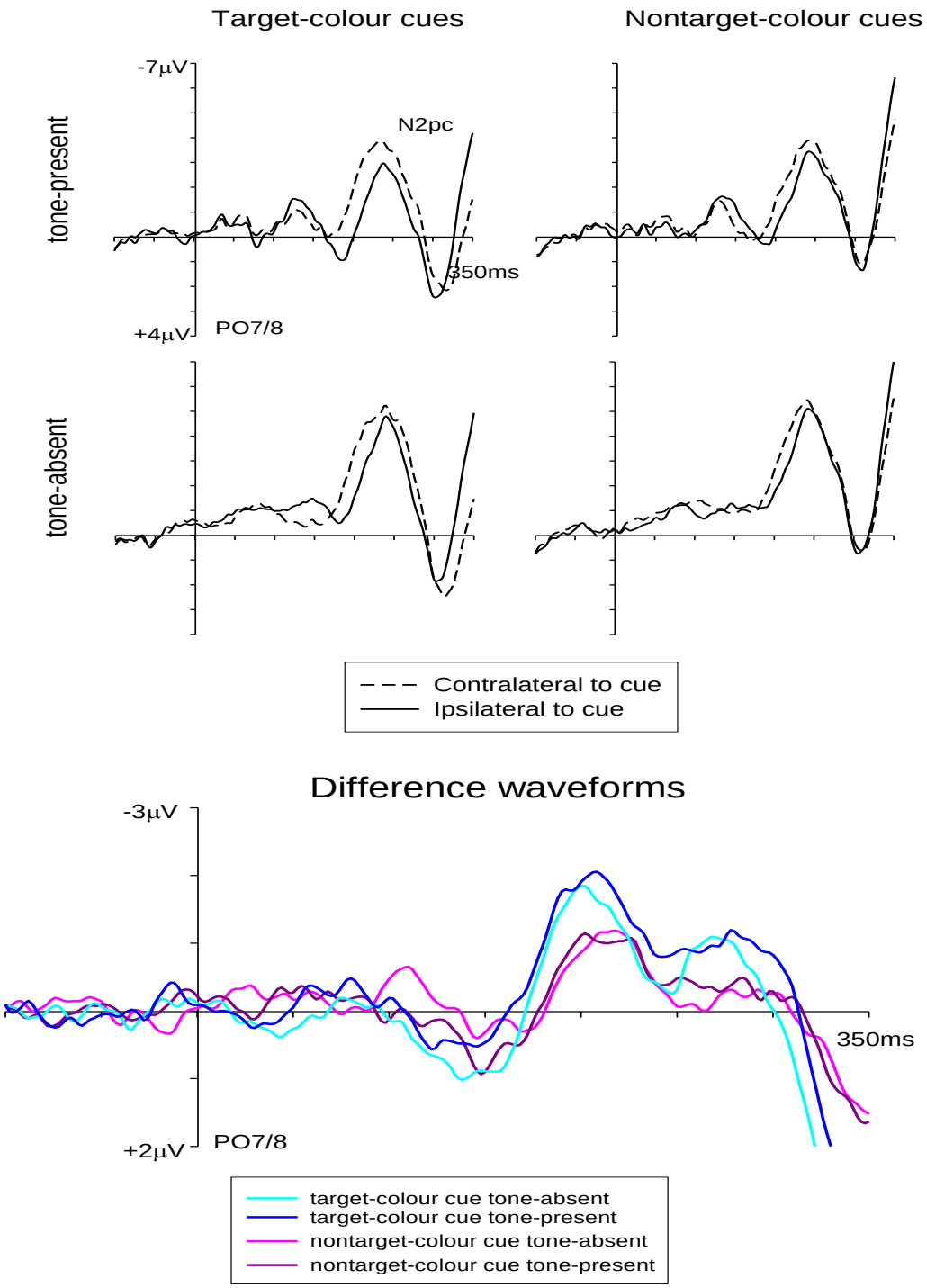

Figure 4.4. *Top panel: Grand averaged visual ERPs obtained in Experiment 6 at electrodes PO7/PO8 contralateral and ipsilateral in response to singleton colour cue arrays, shown separately for target-colour and nontarget-colour cues and for tone-present and tone-absent trials. Bottom panel: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs elicited by singleton colour cues, shown separately for target-colour and nontarget-colour cues, and tone-present and tone-absent trials.*

Discussion

To provide direct evidence for a bottom-up bias in visual selection towards synchronous audiovisual objects, multisensory enhancements in the ability of colour-change cues to capture attention in multi-stimulus contexts were investigated in Experiment 6 with both behavioural and electrophysiological methods. The observed pattern of results was surprising. Despite the fact that the experimental procedures were very similar to ones employed in Experiment 5 (Chapter 3), behavioural spatial cueing effects triggered by heterogeneous colour cues were not larger on tone-present relative to tone-absent trials. Additionally, higher-salience colour singleton cues also did not trigger enhanced RT spatial cueing effects as a function of tone presence.

One possible explanation of these null results is an increase in the perceived task demands associated with EEG recordings. This may have resulted in a stronger top-down inhibition of task-irrelevant tones, thereby preventing multisensory integration from reliably enhancing the ability of visual distractors to capture attention to their location. In other words, the demands associated with a good performance (e.g., additional instructions in respect to eye movements) in an experiment involving EEG recording could have motivated the participants to be inhibiting the tones in Experiment 6 much more strongly than participants taking part in Experiment 5, i.e., a purely behavioural version of Experiment 6. A role of strategic suppression as a factor modulating the audiovisual enhancement of attentional orienting in demanding visual task contexts has been demonstrated previously by Olivers and van der Burg (2008, Experiment 3; for details, see Experiment 3 in Chapter 3), who showed that in such circumstances observers strongly inhibit tones that are task-irrelevant. The inhibition account of the null behavioural results from Experiment 6 is further substantiated by the fact that also the pip-and-pop effect , which is arguably a phenomenon partly driven by audiovisual salience, was shown to be modulated by top-down factors, e.g., with reduced benefits for audiovisual targets found on trials where these targets were presented outside of current focus of attention (van der Burg, Olivers, & Theeuwes, 2012; see also van der Burg et al., 2008a, 2011).

The proposed explanation may account for the behavioural results found in Experiment 6, in which, similarly to other experiments described in Chapters 2 and 3, the location of tones in time was always fully predictable and this point in time was known to be task-irrelevant. Research from the visual domain suggests that, similarly to spatial expectancies, behaviour can be optimised by temporal expectancies, which enable enhanced perceptual processing of events occurring at predicted points in time (Coull, Nobre, & Frith,

2001; Coull & Nobre, 1998; Doherty, Rao, Mesulam, & Nobre, 2005; Griffin, Miniussi, & Nobre, 2001). In one of these studies, Doherty et al. (2005) presented an object moving across a screen that was disappearing behind an occlusion area. Their results demonstrated that perceptual judgements about the reappearing object were facilitated by spatial as well as by temporal expectancies formed about the object by participants, indicating that temporal information can be flexibly used to enhance target processing. It is possible that performance may be optimised by temporal expectancies in a task-dependent fashion not only through top-down enhancements of target processing, but also by strategic inhibition of distractor processing. The surprising absence of audiovisual enlargements of behavioural capture effects triggered by heterogeneous colour cues that was observed in Experiment 6 (cf., Experiment 5 in Chapter 3) may be linked to the fact that cue arrays and the tones that could accompany these cue arrays were always presented at points in time which were known to be task-irrelevant.

In contrast to the behavioural findings, the N2pc results from Experiment 6 revealed reliable multisensory enhancements for one type of colour cue array: Cue-elicited mean N2pc amplitudes were reliably enhanced for audiovisual singleton colour cues. *Prima facie,* these results suggest that at least in some conditions, the N2pc component might be more sensitive to the effects of audiovisual salience than behavioural spatial cueing effects. However, the mixed pattern of electrophysiological results is in fact consistent with the importance of modality-specific salience for stronger bottom-up selection of audiovisual distractors (for details, see General Discussion in Chapter 3). In contexts, where task demands are high, only higher-salience colour singleton distractors will be capable of eliciting a strong bottom-up bias when accompanied by non-visual uninformative signals. The critical role of visual salience for audiovisual enhancements of N2pc mean amplitudes, as revealed by the across-block analysis in Experiment 6, strengthens the conclusions about the importance of unimodal salience for the salience-based audiovisual selection bias that in Chapter 3 were supported solely by behavioural evidence.

However, if unimodal salience was a factor that alone determines the presence of a salience-based bias in visual selection, the N2pc amplitude multisensory enhancements triggered by singleton audiovisual distractors in Experiment 6 should be paralleled by similar enhancements of behavioural spatial cueing effects. This seemingly contradictory pattern of behavioural and electrophysiological findings can be explained by the two-stage account of selection recently proposed by Kiss et al. (2013). Kiss and colleagues (2013) argued that while spatial cueing effects are more indicative of maintenance of attentional focus at the location of the cue until the moment of target stimulus presentation, the N2pc

component reflects an initial stage of attentional selection, where the activity is computed in parallel for each location in external space on a hypothetical salience map (Wolfe, 1994, 2007) on the basis of which attention is allocated. In line with this account, the enhancement of N2pc amplitudes in response to audiovisual singleton colour cues indicate that multisensory integration can enhance the bottom-up bias in visual selection towards irrelevant audiovisual objects in contexts where task demands prevent prolonged engagement of attention in predictable and task-irrelevant location of such salient cues.

In support of the importance of temporal attention for the selection stage reflected by the behavioural spatial cueing effects, pattern of results similar to the current one have been found previously for purely visual salience. Eimer and Kiss (2010; Experiment 2) demonstrated that, in a context encouraging a stringent colour-specific top-down task set, nontarget-colour cues presented at a known-to-be-irrelevant point in time do not trigger reliable behavioural cueing effects. Importantly, the same cues elicited a weak but reliable N2pc component, which suggested that task-irrelevant salient colour singletons captured attention on a subset of trials on which attentional control was reduced. However, it is unclear why the significant N2pc components to the nontarget-colour cues were not paralleled by reliable behavioural cueing effects. Consistent with the two-stage selection account (Kiss et al., 2013), it could be argued that temporal attention can prevent salient visual, as well as audiovisual, distractors from controlling the sustained maintenance of attentional focus in a location of an object in visual space (indexed by spatial cueing effects). In contrast, the earlier selection stage (indexed by the N2pc component), where visual or audiovisual salience provides distractors with a competitive bias, might not be affected by this mechanism. In this context, the findings from Experiment 6 extend the results obtained for temporally predictable purely visual distractors: Even in task contexts where top-down inhibition associated with increased task demands prevents the sustained maintenance of an attentional focus at the location of irrelevant distractors, thus eliminating behavioural cueing effects to subsequent targets, audiovisual salience can further enhance the ability of salient visual distractors to initially trigger stronger attentional capture when accompanied by irrelevant tones.

The second important aim of Experiment 6 was to establish the neural mechanism by which multisensory integration creates a bottom-up bias in spatial selection of visual objects that are paired with non-visual signals. The ERP data demonstrated that audiovisual cues elicited larger N2pc components as compared to visual cues, with no visible differences in N2pc onset latency. This pattern of results indicates that audiovisual synchrony creates a bias in visual object selection (Desimone & Duncan, 1995) by an enhancement of neural

responses triggered by visual distractors, in which it bears resemblance to the neural mechanism responsible for multisensory enhancements of multi-unit and field potential activity in low-level cortices of awake macaque monkeys reported in the literature (e.g., Lakatos et al., 2005, 2007, but see Cappe, Thut, Romei, & Murray, 2010, for evidence that bimodal stimulation recruits different areas in the human brain than unimodal stimulation). The implications of these findings for the nature of the neural mechanism underlying the salience-based audiovisual bias in visual selection will be discussed in more detail in the General Discussion.

To summarise, the aim of Experiment 6 was to provide further evidence for the role of audiovisual salience in enhancing visual object selection by employing the N2pc component as a more direct index of visual attentional capture in multi-stimulus contexts. N2pc amplitude enhancements were found for singleton colour cues (but not for heterogeneous colour cues) that were accompanied by task-irrelevant tones, consistent with the importance of within-modal salience for salience-based multisensory enhancements of attentional capture. In order to replicate and extend the N2pc amplitude modulations by audiovisual salience, in Experiment 7 participants were searching for targets defined by conjunctions of specific visual and auditory features.

Experiment 7. A bottom-up selection bias towards irrelevant audiovisual objects in audiovisual search contexts

Introduction

An insight into whether multisensory integration can bias attentional object selection via a genuine salience-based mechanism can only be provided by tasks in which target and nontarget stimuli are clearly defined. This necessity was recognised by van der Burg et al. (2011) who reported the first ERP study that focused on the audiovisual modulations of the N2pc component triggered by target and nontarget visual stimuli. However, the unusually short time window (cf., Eimer et al., 2009; Hickey et al., 2008; Luck & Hillyard, 1994a) in which reliable N2pc amplitude enhancements were observed warrants care when treating these results as convincing evidence for a bottom-up bias created by multisensory integration in visual selection. The pattern of results found in Experiment 6, in which a simple cueing design (Folk et al., 1992) was employed, was consistent with the important role of visual salience in the presence of bottom-up bias towards audiovisual objects in visual selection. Experiment 7 was designed to replicate and extend the salience-based audiovisual enhancements of the N2pc component, by assessing whether similar enhancements can be found also in task contexts in which visuo-spatial attention is controlled by an attentional template defined by a conjunction of specific visual and auditory features.

All the existing studies that have investigated the role of audiovisual synchrony in creating a bottom-up bias in spatial selection towards audiovisual objects employed tasks in which targets were defined by a single visual feature, such as orientation or colour change (Ngo & Spence, 2010; van der Burg et al., 2008a, 2008b, 2011, 2012). In contrast, in real-life environments observers frequently search for objects, i.e., stimuli that are defined by multiple features; and in many situations these features are coded by different modalities. For example, when we want to localise our mobile phone while it is ringing, the search template guiding our search will be defined by both visual ('small', 'black', rectangular') and auditory (a specific ringtone) features. The biased competition model (Desimone & Duncan, 1995; Duncan et al.,1997; Duncan, 2006) argues that the most important characteristic of the top-down mechanisms biasing visual selection is their flexibility in the

control of selection in accordance with the demands that are created by the task-at-hand. This suggests that, in principle, attention can be biased towards bimodal objects also in a goal-driven manner, i.e., towards objects that share both visual and auditory features of an audiovisually defined target (for details, see Chapter 5). Providing evidence for modulation of visual selection by audiovisual salience in contexts where attention is controlled by target templates that are defined across modalities would allow us to extend the conclusions concerning the important role of multisensory integration as a mechanism creating a bottom-up bias in object selection to real-life contexts.

In order to encourage subjects to search for targets defined by a conjunction of specific visual and auditory features, the majority of trials in Experiment 7 were nontarget trials on which the visual or audiovisual stimuli presented at the onset of the search array matched only one, but not the other target-defining feature (see Figure 4.5). For example, for subjects searching for targets defined as blue bars accompanied by high-pitch tones (V+A+), nontarget trials would include a blue bar accompanied by a low-pitch tone (V+A-), a blue bar presented without a tone (V+), a red bar accompanied by a high-pitch tone (V-A+), a red bar accompanied by a low-pitch tone (V-A-) or a red bar presented without a tone (V-). As temporal attention seems to strongly reduce ERP and behavioural capture effects, a more optimal context to study the bottom-up audiovisual bias in visual object selection was employed in Experiment 7. Auditory stimuli were now presented synchronously with a task-relevant visual search array. This should eliminate effects of temporal attention, as the time at which audiovisual events were presented was now always task-relevant. Critical for the aim of Experiment 7 was the comparison of the N2pc amplitudes elicited by target-colour and nontarget-colour distractor bars in the search array in trials in which these distractor bars were presented without tones (V+ and V- trials, respectively) versus trials in which these bars were accompanied by a nontarget-pitch tone (V+A- and V-A- trials, respectively). In line with the converging evidence for the bottom-up nature of the mechanism by which audiovisual synchrony between irrelevant stimuli can control visuo-spatial attention (for reviews, see Driver & Noesselt, 2008; Kayser & Logothetis, 2007; Koelewijn et al., 2010), it was predicted that even in a context where subjects are searching for a specific colour-pitch conjunction, reliable N2pc amplitude enhancements should be found for both target-colour and nontarget-colour bars accompanied by tones characterised by nontarget pitch.

Method

Participants

Thirteen paid volunteers took part in the study. One participant was excluded due to excessive EEG activity in the alpha band. The remaining twelve participants (mean age 28.5 years, age range 22–38 years; 1 left-handed; 5 males) had normal or corrected vision, and gave informed consent to participate in the study.

Stimuli and apparatus

Visual stimuli were presented at a viewing distance of 100 cm on a 22" LCD monitor (Samsung wide SyncMaster 2233; 100 Hz refresh rate) against a black background. In contrast to Experiment 6, only a search array (50 ms duration) was presented during each trial, followed by an intertrial interval (1450 ms). The search array contained a circular array of six horizontal or vertical bars (1.1° x 0.3°) at an angular distance of 4.1° from a central fixation point (see Figure 4.5), with bar orientation for each position chosen randomly for each trial. On every trial, one of these bars was a colour singleton that could match the target colour (blue or red, varied across subjects; CIE x/y chromaticity coordinates .161/.128 and .621/.128, respectively), surrounded by five uniformly grey (.308/.345) distracter bars. All grey, blue, and red stimuli in the search displays were equiluminant (~11 cd/m^2). The coloured bars appeared equiprobably and randomly at one of the four lateral locations, but never at the top or the bottom location in the array (see Figure 4.5). On some trials, the search arrays were accompanied by an auditory stimulus which was a pure sine-wave tone (50 ms duration; 65 dB SPL; high-pitch: 2000 Hz; low-pitch: 300 Hz) presented concurrently with the search array onset from a loudspeaker located centrally behind the monitor.

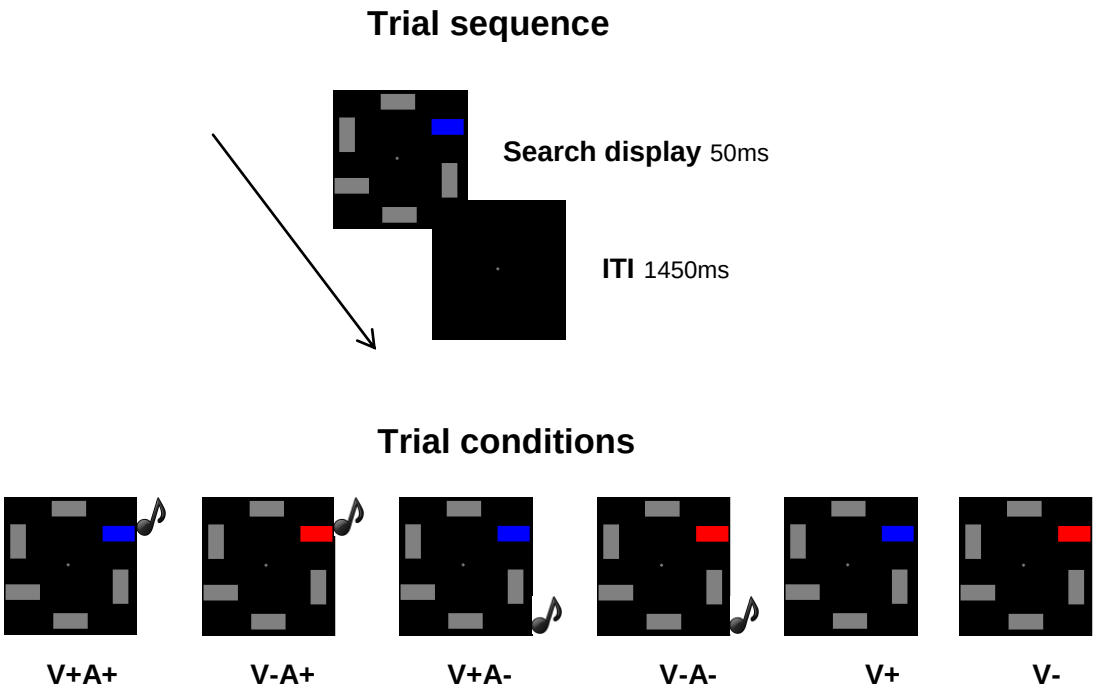

Figure 4.5. *The trial sequence and stimulus setup used in Experiment 7. On each trial, a search array (50 ms duration) contained a colour singleton bar that could match the colour of the audiovisual target or had another, nontarget colour. On some trials, the search array was accompanied by a tone (50 ms duration) that could match the pitch of the audiovisual target or a different, nontarget pitch. All types of trial were equally likely. The example depicts a condition in which the audiovisual target (V+A+) was defined as a blue bar accompanied by a high-pitch tone.*

Procedure and design

Participants were instructed to search for a target bar that was defined by a specific colour-sound combination (e.g., a blue bar accompanied by a high-pitch tone), and to respond to the orientation of this bar by pressing one of two vertically aligned response keys, while refraining from responding on nontarget trials. Target trials were defined as trials on which the singleton colour bar in the search array matched the target colour and was accompanied by a tone that matched the target pitch (V+A+). Nontarget trials were defined as trials on which either one or both of these target-defining audiovisual features were absent (see Figure 4.5). Critical nontarget trials were two purely visual trial types on which the singleton bars that matched the target colour or had another nontarget colour were presented (V+ and V- trials; e.g., a blue bar and a red bar, respectively, presented without tones), and trials on

which these two types of a singleton colour bar were accompanied by a nontarget-pitch tone (V+A- and V-A- trials; e.g., a blue bar and a red bar, respectively, accompanied by a low-pitch tone). The fifth nontarget trial type was defined as a concurrent presentation of a nontarget-colour singleton bar and a tone with a target-matching pitch (V-A+; e.g., a red bar accompanied by a high-pitch tone). This particular type of nontarget trial was included in order to encourage participants to search for the audiovisual target on the basis of both auditory and visual features. All six trial types (target trials and the nontarget trials) were presented with equal probability and in random order. Thus, there was an equal number of target-pitch, nontarget-pitch and no-tone trials, resulting in tones now being presented on 2/3 of all trials. The assignment of target and nontarget colours, as well as of target and nontarget tone frequencies, remained constant for each participant and was counterbalanced across participants. Participants were instructed to respond to the orientation (vertical or horizontal) of a target-colour bar that was accompanied by a target-pitch tone by pressing one of two vertically aligned buttons. Vertical and horizontal responses were mapped onto the top and bottom button, respectively, and the hand-key mapping (e.g., 'top/bottom button - right/left index finger') was counterbalanced within subjects. Four blocks were performed for each of two hand-key mappings, and each sequence of four blocks was preceded by two training blocks. Each block included 96 trials (16 target trials and 16 trials for each of five nontarget trial types), resulting in 768 experimental trials in total.

EEG recording and data analysis

The EEG recording and pre-processing procedures were similar to the ones employed in Experiment 6, with a few notable exceptions. The EEG signal was now time-locked to the onset of the colour bar stimuli, epoched and averaged for the 500 ms interval following the search array onset and relative to a 100 ms pre-array baseline. Averages were analysed only for the four critical nontarget trial types, i.e., trials where target-colour and nontarget-colour singleton bars appeared in the left and right hemifield, with and without the nontarget-pitch tone. A repeated-measures ANOVA for the factors colour-bar type (target vs. nontarget colour), tone presence (tone present vs. absent) and contralaterality (contralateral vs. ipsilateral to the side of presentation of the colour bar) was employed to analyse the mean N2pc amplitudes measured between 190 ms and 290 ms after search array onset at lateral posterior electrodes PO7 and PO8.

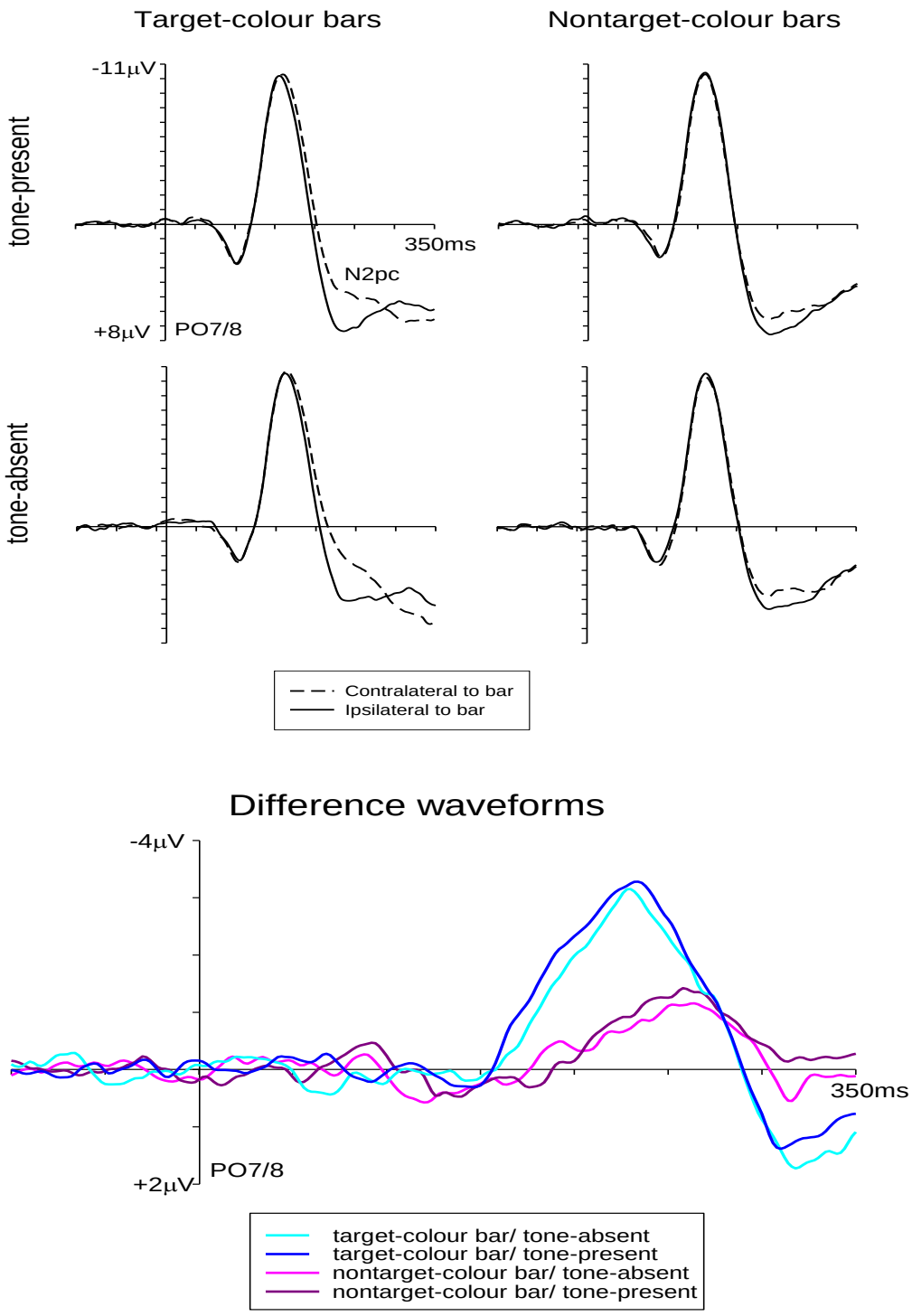

Figure 4.6. *Top panel: Grand-average ERPs measured in Experiment 7 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of target-colour and nontarget-colour singleton bars, separately for tone-present and tone-absent trials. Bottom panel: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the two colour-bar types, and for tone-present and tone-absent trials.*

Results

Only responses from trials with correct responses between 200 ms and 1000 ms were included in the analysis, resulting in an exclusion of less than 1 % of all trials. Target stimuli were missed on less than 1 % of all trials. Participants responded correctly with an average response latency of 594 ms (SEM = 1.5 ms), and made errors on average on 4.8 % (SEM = .06 %) of all trials. False Alarms occurred on .75 % of all trials (SEM = .2 %) and thus were not analysed further. Figure 4.6 (top panels) shows ERPs triggered in response to search arrays at electrodes PO7/8 contralateral and ipsilateral to the side of the target-colour and nontarget-colour singleton bars, separately for trials on which they were presented with and without the nontarget-pitch tone. Similarly to blocks with singleton colour cues in Experiment 6, the N2pc component triggered by the colour singleton bars was enhanced on trials on which these bars were accompanied by a task-irrelevant nontarget-pitch tone, what is more clearly visible in the difference waveforms in Figure 4.6 (bottom panel) computed by subtracting ipsilateral ERPs from the contralateral ERPs.

The reliability of these enhancements was confirmed by statistical analyses carried out on N2pc mean amplitudes. Overall, ERPs were more negative in the N2 time window for target-colour relative to nontarget-colour singleton bars, as revealed by a main effect of colour-bar type, $F(1,11) = 10.12$, $p < .01$, $\eta_p^2 = .48$. ERP amplitudes in the chosen time window were also more positive on trials on which colour bars were accompanied by nontarget-pitch tones compared to tone-absent trials, evidenced by a main effect of tone presence, $F(1,11) = 15.07$, $p < .01$, $\eta_p^2 = .58$. More importantly, the results showed a main effect of contralaterality, $F(1,11) = 49.12$, $p < .001$, $\eta_p^2 = .82$, that was modulated by colour-bar type, $F(1,11) = 19.93$, $p < .001$, $\eta_p^2 = .64$, indicating, as expected, that N2pc amplitudes differed between target-colour and nontarget-colour bars. Pair-wise comparisons revealed that reliable N2pc components were triggered not only on trials on which singleton bars matched the colour of the audiovisual target, $F(1,11) = 41.93$, $p < .001$, $\eta_p^2 = .79$, but also on trials with nontarget-colour singleton bars, $F(1,11) = 29.28$, $p < .001$, $\eta_p^2 = .73$. Critically, a two-way tone presence x contralaterality interaction, $F(1,11) = 4.89$, $p < .05$, $\eta_p^2 = .31$, was also found, demonstrating that N2pc components elicited by colour singleton bars were reliably enhanced on trials on which these bars were accompanied by task-irrelevant nontarget-pitch tones. Follow-up analyses provided evidence that reliable N2pc components were presented on tone-present trials, $F(1,11) = 48.47$, $p < .001$, $\eta_p^2 = .82$, as well as on tone-absent trials, $F(1,11) = 46.68$, $p < .001$, $\eta_p^2 = .81$. A lack of a three-way interaction

between colour-bar type, tone presence and contralaterality, $F < 1$, indicated that these enhancements were similar in size for target-colour and nontarget-colour singleton bars.

Discussion

In Experiment 7, participants were searching for targets defined by a conjunction of specific visual and auditory features, rather than by a single visual feature, as in Experiment 6. In spite of this difference, the N2pc components triggered in Experiment 7 in response to singleton colour bars were enhanced by the presence of a task-set irrelevant tone, and these audiovisually-induced enhancements were similar for colour bars that matched the target colour and those that did not. The design of Experiment 7 prevented any behavioural measurement of the multisensory enhancement of visual selection from being employed. However, the critical importance of Experiment 7 was that it extended the findings from Experiment 6 to more ecologically valid task contexts. Namely, the N2pc enhancements found in Experiment 7 demonstrated that selection of visual objects in space can be biased towards irrelevant audiovisual objects even during search for objects defined by specific visual and auditory features, i.e., in task contexts where attentional control settings are configured for features in the visual as well as in the auditory modality. Importantly, in contrast to van der Burg et al. (2011), but similarly to Experiment 6, the N2pc components triggered in response to singleton colour bars were enhanced by multisensory integration in a large time window spanning 100 ms, which provided converging evidence for the role of audiovisual salience as a source of substantial modulation of visual object selection.

Importantly, conclusions with respect to the task-set independent nature of the synchrony-induced N2pc modulations observed in Experiment 7 need to be qualified. On the one hand, in spite of the fact that nontarget-pitch tones were by definition task-irrelevant, participants still had to process them because auditory target-nontarget discriminations were required on each trial to determine whether a specific target-defining combination of visual and auditory features was present, and in order to decide whether to execute a response on a given trial. In other words, nontarget tones were not strictly task-irrelevant in Experiment 7, and this fact could in principle have contributed to the amplitude enhancement found on audiovisual versus visual trials. On the other hand, the type of experimental setup employed in Experiment 7 provides an important validation of the audiovisual salience-based bias in object selection in more ecologically valid contexts. Namely, the findings from the present study demonstrated that even in circumstances in which the search target is defined across modalities, and focal attention needs to be deployed to the visual as well as auditory features

of a bimodal event in order to categorise it as a target or a nontarget stimulus (see Eimer et al., 2002; Tellinghuisen & Nowak, 2003, for evidence of partially independent processing resources in vision and audition), synchrony-induced modulations of visual object selection indexed by the N2pc component are similar for target- and nontarget-colour bars. These results strengthen the previous conclusions about the salience-based nature of the bias created by audiovisual salience in visual object selection by demonstrating that even in contexts where tones and visual stimuli have to be processed up to the level of their identity, the competitive bias provided to a visual stimulus by audiovisual salience is not modulated by the identity of this visual stimulus.

Conclusions from Chapters 2–4

To cope with the high processing demands that are typical for perceptually cluttered real-life environments, the brain has developed mechanisms that automatically facilitate processing of events for which crucial information as, for example, location in time, is provided concurrently by two sensory modalities, rather than just a single modality. However, existing studies failed to show convincingly that in multi-stimulus contexts, audiovisual synchrony can provide visual objects accompanied by irrelevant tones with a competitive advantage that is driven purely by a bottom-up mechanism (see van der Burg et al., 2008a, 2008b, 2011). The experiments that are reported in Chapters 2 to 4 have revealed the neural and cognitive mechanisms that underlie the bottom-up bias that can be created in visual object selection by multisensory integration, and also brought important novel insights into the factors that modulate this phenomenon.

By means of the spatial cueing paradigm introduced by Folk and colleagues (Folk et al., 1992), it was demonstrated that multisensory integration can enhance the ability of irrelevant visual objects paired with irrelevant non-visual signals to attract shifts of visuo-spatial attention to their location in multi-stimulus contexts. Beside their enhanced control over the involuntary orienting attentional system, synchronous bimodal objects were also shown to trigger stronger spatially selective brain responses as compared to unimodal objects. Importantly, in all the experiments reported in this part of the present thesis, the magnitude of audiovisually induced enhancements of visual attentional capture, whether indexed by RTs spatial cueing effects or by the N2pc component, was similar for the visual distractors that shared the target-defining feature and those that did not. These results

support the account that proposes that synchrony-driven enhancements of early neural responses to visual events in low-level sensory cortices (for reviews, see Cappe et al., 2010; Driver & Noesselt, 2008; Kayser & Logothetis, 2007) create salient synchronous multimodal objects, which are preferentially selected at later stages of cortical processing, during which these multimodal objects compete with purely visual objects for control over perception, memory encoding and action (Desimone & Duncan, 1995; Duncan et al., 1997).

Further evidence for the salience-based nature of the mechanisms responsible for the preferential selection of audiovisual over visual distractors was provided by the experiments reported in Chapters 2 to 4 that focused on the factors that modulate this form of bottom-up bias. In the existing literature, only abruptness (van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; Vroomen & de Gelder, 2000) and the stimulus delivery rate of the auditory signal (Kösem & Van Wassenhove, 2012), i.e., the bottom-up factors that were controlled in all of the current experiments, were shown to determine the effectiveness of audiovisual synchrony in creating a bottom-up bias in visual object selection. In the present experiments, within-modal salience, either visual or auditory, was repeatedly shown to modulate the multisensory enhancements of visual attention capture, as indexed by behavioural (across-experiment comparisons: Experiment 1 vs. Experiment 4; Experiment 3 vs. Experiment 5) or electrophysiological (across-block comparison in Experiment 6) measures. These results provided novel evidence for a threshold-like nature of the mechanisms responsible for creating a reliably stronger bottom-up bias in attentional selection for audiovisual relative to visual objects.

With respect to top-down factors, size of the attentional window has been the only factor known to modulate the magnitude of selection enhancements that are based on audiovisual salience (van der Burg et al., 2012). In contrast to within-modal salience, in the experiments reported in this part of the present thesis, the search strategy encouraged by task instructions did not seem to modulate the presence or the magnitude of the audiovisual enhancements of visual object selection: For pairings of lower-salience heterogeneous colour cues with lower-intensity tones, no evidence of enhanced attentional capture effects was found, irrespective of whether a singleton-detection or a feature-search mode (Experiment 3 vs. Experiment 4; Bacon & Egeth, 1994) was adopted by the participants. Additionally, higher-salience colour singleton distractors triggered enhanced N2pc amplitudes irrespective of whether the target was defined by a single feature (Experiment 6) or by a conjunction of visual and auditory features (Experiment 7). What is important, the findings from Experiment 6 suggested that temporal attention might be an important top-down factor that determines presence of salience-based multisensory enhancements of

attention capture in vision: In this experiment, temporal attention prevented both higher- and lower-salience audiovisual distractors presented at known-to-be-irrelevant locations from attracting visuo-spatial attention more strongly than unimodal visual distractors. However, the activity on the hypothetical salience map responsible for the allocation of attention to objects in space (Kiss et al., 2013; Wolfe, 1994, 2007), reflected by the N2pc component, seems not to be subject to this form of attentional control, as higher within-modal salience enabled the colour singleton cues to trigger reliably enhanced N2pc amplitudes when accompanied by tones. It is possible that in contexts where no temporal expectancies are present with respect to the occurrence of audiovisual distractors, as is often the case in real-life environments, more robust salience-based enhancements of visual object selection might be observed, as hinted on by research from the visual domain (Eimer et al., 2009; Lamy et al., 2003, 2004).

Overall, the behavioural and electrophysiological findings reported in Chapters 2 through to 4 reveal a neuro-cognitive mechanism of a rather different nature than the multisensory enhancements of neural responses and behavioural orienting that were originally described by Stein and colleagues (see Stein & Meredith, 1993). The latter mechanisms, likely mediated by the SC, are contingent on spatiotemporal alignment of events from different modalities and are associated with a close spatial register of receptive fields within multisensory neurons in the deep layers of SC for stimuli from different modalities. At the neural level, this type of multisensory enhancements is reflected by a combination of amplitude and onset latency effects (e.g., Bell et al., 2005), indicative of multisensory integration facilitating orienting behaviour via a combination of magnitude- and speed-related effects. Critically, these enhancements of responses to bimodal stimuli, when compared to the sum of responses elicited by both unimodal stimuli in isolation, are observed most frequently in contexts where one of these stimuli could be not registered if presented alone (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Holmes & Spence, 2005; Stein et al., 1988; Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009). Thus, this particular multisensory mechanism facilitates processing of weak peripheral signals from different modalities that might indicate a single event or object, which suggests specialisation of this mechanism in detection of faint but potentially important events presented outside of the attentional focus (Stein, 1998; Stein, Stanford, & Rowland, 2009; Stein & Stanford, 2008).

There are numerous contrasts between this multisensory mechanism and the mechanisms that were investigated in this part of the present thesis. First, mere temporal synchrony between signals from different modalities will be often sufficient (Stein et al.,

1996; Talsma et al., 2010) to provide visual objects accompanied by non-visual signals with a competitive advantage in neural networks involved in spatially specific perceptual encoding, behaviour control, and representation in the short-term memory (Desimone & Duncan, 1995; Duncan et al., 1997). Second, this form of a salience-based audiovisual bias in visual object selection seems to be contingent on modality-specific salience, where multisensory enhancements of attentional capture are absent or suppressed, rather than enhanced, in contexts where one of the stimuli is not of sufficiently high salience. Finally, this form of audiovisual bias in attentional selection may even reflect an innate (see Spector & Maurer, 2009, for a review of evidence for a superabundance of cross-modal connections in the infant brain) processing bias towards bimodal stimuli that is present in multiple brain networks involved in visual processing, arguably due to a potential stronger behavioural relevance of bimodal relative to unimodal events: There is a growing body of evidence that over early development, attention and learning is typically biased towards redundantly defined aspects of the external environment as this implies their importance (for a review, see Bahrick, Lickliter, & Flom, 2004). In contrast, the multisensory enhancement of attention orienting is a mechanism activated only in very specific contexts, where otherwise faint or altogether ineffective peripheral signals are redundantly defined in both space and time (Bolognini et al., 2005; Ho et al., 2009; Santangelo & Spence, 2007a; Stein, 1998), and seems to be a processing *capacity* that a brain develops only after it has accumulated substantial within-modal experience (Jiang, Wallace, Jiang, Vaughan, & Stein, 2001; Wallace, Carriere, Perrault, Vaughan, & Stein, 2006; Wallace, 2004).

In the present thesis, the biased competition model (BCM; Desimone & Duncan, 1995; Duncan et al., 1997, 2006) was employed as a theoretical framework to aid understanding of the mechanisms of selective processing of visual objects which bottom-up salience is increased by audiovisual synchrony (see also Bishop, 2008, for the usage of BCM as a framework to explain selective processing of threat-related stimuli). The results reported across Experiments 1 to 7 provide converging evidence for the assumptions of the BCM that were formulated originally with respect to purely visual processing. Consistent with this model, current behavioural and ERP findings suggest that in multi-stimulus contexts, even entirely irrelevant visual objects that are accompanied by non-visual signals have an enhanced ability to control the location of visuo-spatial attention, and that this enhancement is accompanied by a competitive advantage for such salient bimodal objects at the stage concerned with selective perceptual processing, where both are the result of effortless multisensory integration taking place at early levels of cortical processing. Based on the evidence from the area of visual attention (cf., Eimer et al., 2009; Lamy et al., 2004),

it can be assumed that in contexts where task-irrelevant visual objects do not appear at known irrelevant locations in time, a stronger visual selection bias towards visual objects accompanied by non-visual signals would be observed with both the ERP and behavioural markers of attentional capture. This likely pattern would be in line with the tenet of BCM, according to which outcome of the stimulus competition is integrated across different neural visual systems, i.e., those involved in selective perceptual processing and those responsible for behavioural control.

Overall, the current findings seem to support the conceptualisation of selective attention proposed by BCM according to which selective attention is an emergent 'state' of the cognitive system (Desimone & Duncan, 1995), which results from focusing the processing across multiple neural networks onto a single object. What is worth noting, BCM also seems to provide an explanation for the critical importance of within-modal salience for the presence of an audiovisually induced bottom-up selection bias: In heterogeneous cue displays, colour-change cues may have not been sufficiently salient to win the competition on each trial, which may have effectively decreased the number of trials on which the concurrent tone was successfully integrated with these cues and therefore resulted in a stronger bottom-up bias. In contrast, in displays where cues were feature singletons, competition between multiple simultaneous objects should not occur (see Beck & Kastner, 2009, for a review of the neuroimaging evidence in support of the role of within-modal salience in resolving the competition in visual brain areas), thus allowing for the accompanying tones to be successfully associated with the colour cues on every trial, thereby creating a salient bimodal object. In other words, increased within-modal competition may be another crucial factor that determines the presence of an audiovisual bottom-up bias in visual object selection (see also Sanabria, Soto-Faraco, Chan, & Spence, 2005; Sanabria, Soto-Faraco, & Spence, 2004).

At the same time, the current findings provide an important extension of the BCM model by demonstrating a novel form of bottom-up bias, one that contrasts with the biases associated with purely visual salient distractors in that it will often affect attentional object selection independently of top-down task set. However, the ability of bimodal distractors to create a reliable bottom-up bias during both early-stage and the late-stage selection might be under the control of temporal attention. In this respect bimodal objects might be similar to salient visual distractors (cf., Eimer et al., 2009, 2010; Theeuwes et al. , 2000). Overall, the results reported across Chapters 2 to 4 cannot be easily accommodated within the models that propose either that selection in vision is always contingent on top-down control settings (Folk et al., 1992, 1998), or that the early selection stage is controlled solely by bottom-up

salience (Theeuwes et al., 2000; Theeuwes, 2010). Similarly to BCM, the current version of the Guided Search model, another major model of visual attention (Wolfe et al., 1989; Wolfe, 1994, 2007), also recognises the importance of top-down and bottom-up mechanisms in the control of visual selection, which mechanisms, it argues, jointly influence the activation in locations in external space that are represented in a single hypothetical saliency map responsible for allocation of visuo-spatial attention. However, because this model argues for only a single stage at which attentional selection occurs, its cannot explain why colour singleton distractors trigger stronger spatially-selective perceptual processing when accompanied by tone even when this effect is not accompanied by stronger behavioural capture effects.

To sum up, the results reported throughout Chapters 2 to 4, indicative of a new form of a bottom-up bias in visual object selection, highlight the importance of multisensory research for the theories of selective attention, where conclusions are frequently limited to purely visual processing. Additionally, these findings further illustrate the flexibility of the BCM in explaining different types of bottom-up bias outside of the ones typically considered by visual attention researchers.

Chapter 5. Top-down control of audiovisual search by bimodal search templates

Chapter 5 describes four experiments that investigated whether multisensory integration can bias attentional selection towards audiovisual objects via a top-down, goal-based mechanism. In real-life environments, only a small subset of the information that is entering our senses at each point in time can be fully processed, and, thus, external objects constantly compete with each other for access to perception, memory and control over behaviour. In such circumstances, effective behaviour is critically dependent on mechanisms that bias competition in these visual systems in favour of objects important to behavioural goals (Desimone & Duncan, 1995; Duncan et al., 1997). Goal-based preferential processing is instantiated and controlled by working memory representations or 'attentional templates' of currently goal-relevant features of the environment (e.g., Duncan & Humphreys, 1989; Olivers & Eimer, 2011; Olivers, Peters, Houtkamp, & Roelfsema, 2011; but see Carlisle, Arita, Pardo, & Woodman, 2011; Olivers, 2011; Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006; for evidence that top-down selection bias can be controlled by long-term representations). Most investigations into the nature of attentional templates have focused on the visual modality, and on tasks where task-relevant stimuli are defined in terms of one specific feature or feature dimension (e.g., 'red' or 'targets defined by colour'; e.g., Bacon & Egeth, 1994; Eimer & Kiss, 2008; Eimer et al., 2009; Folk et al., 1992; Lamy et al., 2004).

However, in naturalistic environments attentional selectivity is rarely directed towards single elementary perceptual features (e.g., 'red' or 'round'). In the real world, we typically search for objects that are defined by a conjunction of features from different dimensions (e.g., search for a black, small, and rectangular mobile phone). Furthermore, in real-life environments we frequently use simultaneous cues from different sensory modalities to locate target objects. For example, when we want to find our misplaced mobile phone while it is ringing, the attentional template guiding our search will include both visual (colour, size, shape) as well as auditory features (the pitch or melody of the ring-tone). On the one hand, in multisensory environments spatial attention might be typically dominated by visual representations of space (Eardley & van Velzen, 2011; Eimer et al., 2002; Welch & Warren, 1980). This would suggest that that purely visual objects could capture attention reliably even during search for a target defined across visual and auditory modality. On the other hand, in line with the models that argue

for the dominating role of top-down factors in the control of visual selective attention (Desimone & Duncan, 1995; Duncan et al., 1997; Wolfe, 1994, 2007) it can be assumed that attentional selection should be effectively guided towards target objects even when these targets are defined across modalities. The idea that attention might be preferentially deployed to audiovisual relative to visual objects via a goal-based mechanism, i.e., because they match both features of a bimodal target, is supported also by the body of research that highlights the fact that some forms of multisensory integration take place after the initial selection of each signal within the modality-specific cortices (see Sections 1.2.3 and 1.3.3, for more details).

Even in the visual domain, the issue of how attentional templates are structured in contexts where targets are defined by multiple features has only been investigated very recently (Irons & Remington, 2013; Kiss et al., 2013; Leblanc, Prime, & Jolicoeur, 2008). There are two possible ways in which task-relevant features can be represented in attentional templates in such contexts. Most current models of visual attention (Treisman & Gelade, 1980; Wolfe, 1994, 2007) assume that visual search is guided independently by separate representations of task-relevant features. This account argues that objects that share only some of the target features (e.g., a paper notebook that matches the colour and shape, but not the size of the mobile phone) will reliably capture attention because separate features guide attention in feature-conjunction tasks. According to the Guided Search model (Wolfe, 1994, 2007), the allocation of attention is controlled by a central spatiotopically organised salience map that receives inputs from anatomically separate and independently operating visual feature channels. Because top-down weighting occurs in an independent manner for each channel, and because inputs contributing to the activity profile on the salience map do not interact with each other, GS supports a stance according to which attentional control is guided via independently represented features. This account is also consistent with the 'task-set contingent attentional capture' hypothesis (Folk et al., 1992), which proposes that presence of a *single* task-relevant feature in an irrelevant object will result in its involuntary selection even in circumstances where that stimulus is presented in locations in space or time that are known to be task-irrelevant (cf., Eimer et al., 2009).

An alternative account argues that in contexts where targets are defined as conjunctions of different features, these features are represented as integrated object representations. This implies that objects which share only some of the target features (e.g., a paper notebook that looks similar to our mobile phone) will fail to capture attention, because single features do not guide attention in feature-conjunction search contexts. This account is supported by the biased competition model (Desimone & Duncan, 1995; Duncan et al., 1997), which proposes that objects act as wholes during competition for processing resources and that integration of separate

features into unified objects often occurs in parallel prior to attentional selection (in contrast to Treisman & Gelade, 1980, who claimed that attention is required to join features into object representations). Importantly, working memory, which is traditionally regarded as the locus of attentional templates, seems to represent features in a form of integrated objects, rather than independent features (e.g., Luck & Vogel, 1997).

The study of task-set contingent attentional capture offers a valuable tool to investigate the internal organisation of attentional templates. If attentional templates represent independently each feature of a currently task-relevant object, any object that matches at least one of these features should attract attention, irrespective of whether other target-defining features are also present. If this account is correct, then, for example, during search for a lost mobile phone, attention would be captured by all objects that match the size, shape, or colour of this phone. Instead, if attention during multi-feature search is guided by fully integrated object representations, individual target features should not affect attentional processes in isolation. Partially matching objects would be unable to capture attention, which will only be attracted by objects that possess all currently task-relevant features. In a recent study aimed at directly testing the two accounts, Kiss and colleagues (2013) demonstrated that in fact both are correct, as they describe two successive stages of top-down control of visual object selection during multi-feature target search. In one of their experiments (Kiss et al., 2013; Experiment 1), participants searched for targets defined by a conjunction of colour and size (e.g., small red bars), and the search arrays were preceded on each trial by cue arrays. Critically, cues were spatially uninformative feature singletons that could match both target-defining features, one of the features, or neither of them. The cues that shared only one of the target-defining visual features (e.g., small blue cues during search for small red targets) captured attention in a task-set contingent fashion, as indicated by the presence of a cue-elicited N2pc component. However, the same cues did not trigger reliable behavioural spatial cueing effects, suggesting that attention was immediately disengaged from the cue location after initial capture due to their nontarget status (for more details, see Experiment 8 in the present chapter).

The research reported in this chapter was motivated by the fact that in real-world environments search is frequently directed towards objects whose relevant features are defined in different sensory modalities (e.g., the mobile phone with its personalised ringtone). Thus, the question of *independent* versus *integrated* representations of target-defining features arises also for search objects defined across modalities. While the study of Kiss et al. (2013) demonstrated that during search for multi-feature visual targets visual selection is guided at least in part by integrated object representations, it provides no evidence that attentional templates are structured

in the same way for targets defined by features from two different modalities. Research investigating the interplay between multisensory processing and selective visual attention has typically focused on spatial synergies in attention across sensory modalities (Eimer et al., 2002; Spence & Driver, 1996; van Velzen et al., 2006) or a bottom-up bias in visual selection towards visual objects accompanied by non-visual signals (Matusz & Eimer, 2011; Olivers & van der Burg, 2008; van der Burg et al., 2008a). It is possible that in contexts where the audiovisual objects match both features of an audiovisually defined target, a top-down, goal-based mechanism might be another source of bias in spatial selection towards audiovisual relative to visual stimuli.

Until now, only two studies have investigated whether spatial selection of a visual object can be facilitated in contexts where the visual object is accompanied by a sound that also characterises the target (Iordanescu et al., 2010; Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008). Iordanesco et al. (2010) instructed participants to search for naturalistic objects (e.g., 'dog') in a six-object array. Saccadic search times were reliably reduced relative to a no-sound condition in contexts where a semantically congruent but spatially uninformative sound (e.g., dog bark) was presented concurrently with the search array. Interestingly, presentation of a sound that was semantically congruent with one of the five distractor objects present in the search array (e.g., clock ticking) did not lead to a delay in saccadic search times relative to the no-sound condition. These results can be explained by a top-down goal-based mechanism whereby spatially-specific processing of a visual target is facilitated by the presence of a redundant semantically congruent non-visual signal (see van der Burg et al., 2008a, for evidence of facilitation of search behaviour by temporal redundancy). This effect is unlikely to be driven by a simple alerting mechanism, as in another study Iordanesco et al. (2008) showed that the presentation of an unrelated sound (e.g., mosquito buzzing) at the onset of the visual search array did not lead to a reduction of saccadic search times.

While demonstrating a new form of cross-modal facilitation of attentional selection in multi-stimulus contexts, the findings of Iordanesco et al. (2008, 2010) do not provide an insight into the nature of attentional templates that guide attentional selection during search for targets defined by a conjunction of features from different modalities. Those findings also do not address whether attention can be controlled by multimodal object templates in contexts where targets are defined by combinations of *arbitrary* features (i.e., not semantically related). In the Kiss et al. (2013) study, targets were arbitrary conjunctions of colour (red or blue) and size (large or small), counterbalanced across subjects, which allowed the authors to demonstrate the effectiveness of top-down guidance of selection by colour-size target templates while preventing the results from

being confounded by pre-existing feature associations (cf., Evans & Treisman, 2010). Thus, a pressing question that remains unanswered is whether attentional selection in multi-feature target search task contexts can be controlled by integrated arbitrary object templates also for targets that are defined across modalities? Research into the nature of attentional templates controlling search for multimodal targets will provide a better understanding of feature-based top-down mechanisms guiding attentional selection in real-life environments, in which targets are frequently defined across modalities.

Analogous to the purely visual case, multimodal attentional templates might be composed of single crossmodal object representations that integrate features from different modalities. According to this account, a distractor that matches all the visual features but not the auditory feature of an audiovisually defined target (e.g., a black, small paper notebook), should fail to capture and hold attention because it would not match fully the bimodal target template. This account is supported by the biased competition model (Desimone & Duncan, 1995; Duncan et al., 1997), which was successfully employed as a theoretical framework to explain the findings indicative of a bottom-up selection bias towards synchronous bimodal objects (see Chapters 2 to 4). According to this model, effective cognitive functioning necessitates a ready adjustment to the demands of the task-at-hand, thus rendering the flexibility in creating representations of task-relevant features the most important characteristic of the top-down control mechanisms (Duncan et al., 1997). In line with the biased competition model, in contexts where targets are defined by conjunctions of features from different modalities, a fully integrated multimodal object template should be constructed to effectively guide attentional selection.

Support for the integrated multimodal object template account is provided by human neuroimaging and animal neurophysiological studies, which showed that areas involved in top-down control of visuo-spatial attention (see Corbetta & Shulman, 2002; McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2003) are also known loci for late-stage multisensory integration. The lateral intraparietal sulcus, an area recruited during the control of attention based on visual locations, features and objects (Assad & Maunsell, 1995; Shulman et al., 2002), is known to contain a multisensory map of space (Andersen, 1997; Andersen, Snyder, & Bradley, 1997; Stricanne, Andersen, & Mazzoni, 1996). The dorsolateral prefrontal cortex, an area critical for working memory, contains neurons that can maintain integrated representations of cross-modal pairings of colour and pitch when they are relevant to the task (Fuster et al., 2000; but see Warden & Miller, 2007, for evidence that object selectivity in PFC might be weak). The superior temporal sulcus, a neural substrate for cross-modal effects in spatial attention (McDonald et al., 2003), is involved in the integration of perceptual and semantic features from different modalities (Taylor

et al., 2006; Werner & Noppeney, 2010). Critically, the right inferior frontal cortex has recently been highlighted as crucial for the integration of unfamiliar artificial sounds and images into multimodal objects, with the posterior superior temporal sulcus (and superior temporal gyrus) involved in the integration of audiovisual pairings that are more familiar and semantically congruent (Hein, Doehrmann, Müller, Kaiser, Muckli, & Naumer, 2007; Naumer, Doehrmann, Müller, Muckli, Kaiser, & Hein, 2009). The evidence for goal-based multisensory integration in areas recruited by attentional control in visual tasks suggests multiple brain loci in which fully integrated object templates could be created and maintained with the purpose to guide attentional object selection flexibly in multimodal contexts (cf., Duncan et al., 1997).

An alternative account might argue that multimodal attentional templates are implemented as anatomically and functionally independent modality-specific representations of target-defining features. If search for cross-modal targets is guided by independent within-modality target templates, a distractor that matches all the visual features of the mobile but is lacking the auditory feature (e.g., a notebook that looks similar to our mobile phone) should capture and hold attention (cf., Kiss et al., 2013) because it would fully match the *visual* target template. Importantly, in search for multi-feature targets, spatial selection might be controlled by integrated object templates in the case of visual combinations, while being controlled by independent within-modality templates for arbitrary audiovisual pairings. The view proposing separate within-modality attentional templates is in line with the superior spatial resolution of the visual system (Welch & Warren, 1980) that results in the dominance of vision in processes concerning object recognition and action in external space (Goodale & Milner, 1992; Mishkin & Ungerleider, 1982). This account would also be supported by older versions of the Baddeley's (1998; see also Baddeley & Hitch, 1974) working memory theory, according to which stimulus representations are maintained in two separate modality-specific 'slave systems' controlled by the central executive, i.e., the visual sketch-pad and the auditory phonological loop (but see Baddeley, 2000, for proposal of an additional, inter-modal component of working memory, i.e., episodic buffer).

Whether attentional selection is controlled by integrated bimodal templates or separate within-modality representations of target features can be assessed by investigating attentional capture effects during search for target objects defined by a combination of features from different modalities. In Experiment 8, behavioural and ERP measures of attentional capture were employed to measure the ability of target-matching colour cues to capture attention across search tasks in which targets were defined by colour only or by a combination of colour and pitch. Experiment 9 was designed to investigate whether more reliable effects of audiovisual templates

on task-set contingent capture by target-matching visual cues can be observed in contexts where the target pitch is more predictive of presence of the bimodal target. The aim of Experiment 10 was to provide an insight into the role of the relative salience of unimodal cues in their ability to capture attention in audiovisual search task contexts. Experiment 11 was designed to assess whether stronger effects of top-down guidance of attentional selection by fully integrated audiovisual object representations is observed in circumstances where the targets are defined on a dimension different than colour, i.e., size.

Experiment 8. Top-down guidance of visual selection by integrated audiovisual search templates

Introduction

To test whether attention is controlled by integrated object representations of individual features in multi-feature search target contexts, Kiss et al. (2013; Experiment 1) used a cueing paradigm (Folk et al., 1992), and measured RT and ERP markers of attentional capture elicited by feature singleton cues that preceded the search array on each trial. Participants were instructed to search for targets defined by a specific conjunction of visual features (e.g., small red bars) presented on half of all trials to make an orientation judgement about them. Cues were spatially uninformative colour/size singletons that could match both (C+S+), one (C+S-, e.g., large red cues; C-S+, e.g., small blue cues) or neither (C-S-; e.g., large blue cues) of the target-defining features. Behavioural spatial cueing effects were triggered by fully matching cues, but not by fully non-matching cues, confirming that attentional capture is task-set contingent (Folk et al., 1992), and suggesting that in multi-feature target search contexts visual attention is controlled in a top-down fashion by integrated target templates. In contrast to behavioural effects, reliable N2pc components were observed in response to both fully and partially matching cues, implying that the presence of a single task-relevant feature suffices for attention to be automatically captured in a task-set contingent fashion to the location of an irrelevant distractor, and indicating attention control by independent feature lists. Kiss et al. (2013) proposed a two-stage model, according to which the critical feature of top-down guidance by feature-conjunction task sets is that attention is initially attracted to distractor possessing task-relevant features (cf., Theeuwes et al., 2000). In the absence of a full match with the target template, an immediate disengagement from the distractor location is triggered, resulting in the absence of behavioural spatial cueing effects for partially matching cues. The question is whether the same two levels of top-down control over visual selection can be found in contexts where targets are defined across modalities.

Whether multimodal attentional templates are anatomically and functionally independent modality-specific representations of target-defining features or single crossmodal object representations that integrate features from different modalities can also be assessed using task-set contingent attentional capture. In Experiment 8, participants were instructed to search for targets that were either defined by a single visual feature or by a combination of a visual and an auditory feature. One of the important insights provided by the study of Kiss et al. (2013) is that

care is warranted when drawing conclusions about top-down control of attentional selection solely on the basis of behavioural measures as they might reflect only one of the stages of attentional object selection in contexts that measure task-set contingent attentional capture. Thus, in Experiment 8 the ability of task-set matching visual singleton cues to capture attention was assessed in unimodal visual and bimodal audiovisual task contexts with both behavioural and ERP measures.

The task procedures employed in Experiment 8 were similar to those used by Kiss et al. (2013) to study attentional control during search for colour-size targets. In the present experiment, on each trial, a spatially uninformative colour singleton cue preceded a search array that included a red or blue colour singleton bar (see Figure 5.1, top panel). On target trials, participants had to discriminate the orientation (horizontal versus vertical) of this bar. In the unimodal 'Colour' task, they had to respond to red bars and ignore blue bars (the target/ nontarget colour assignment was counterbalanced across participants). In two audiovisual tasks, search arrays could be accompanied by synchronous tones, and target trials were defined by a combination of visual and auditory features. The two audiovisual tasks were employed to explore how the presence versus absence of the additional requirement to discriminate tone frequency on each trial affects attentional capture by colour-matching cues. In the 'Colour-Sound' task, 'tone present versus absent' judgements sufficed to distinguish target and nontarget trials. Participants had to respond to target-colour bars that were accompanied by a tone (V+A+ trials; e.g., a blue bar accompanied by high-pitch tone), and to ignore target-colour bars without tones (V+ trials; e.g., a unimodal blue bar), as well as all nontarget-colour bars, regardless of whether they were presented with or without tone (V-A+ and V- trials; e.g., a red bar accompanied by high-pitch tone and a unimodal red bar, respectively). In the 'Colour-Pitch' task, all search arrays were accompanied by a tone, and tone pitch discriminations ('high vs. low') were required. Target trials were defined by a specific colour-pitch combination (V+A+ trials). Nontarget trials were defined as trials with target-colour bars accompanied by nontarget-pitch tones (V+A- trials; e.g., a blue with a low-pitch tone), nontarget-colour bars accompanied by target-pitch tones (V-A+ trials; e.g., a red bar with a high-pitch tone), and nontarget-colour bars accompanied by nontarget-pitch tines (V-A- trials; a red bar with a low-pitch tone). In all three tasks, the colour singleton cue always matched the current target-defining colour.

For the unimodal Colour task, behavioural and electrophysiological attentional capture effects were expected to be similar to those found in previous unimodal visual spatial cueing experiments (e.g., Eimer & Kiss, 2008; Folk et al., 1992). Attentional capture by target-colour matching singleton cues should be reflected by faster RTs to target bars at cued as compared to

uncued locations, and by the presence of N2pc components in response to these singleton cues. The critical question was whether the same effects would also be observed during search for audiovisually defined targets. If this search was guided by strictly modality-specific attentional templates that operate independently for visual and auditory target-defining features, the fact that the colour cues matched the currently task-relevant colour in all three tasks should result in identical attentional capture effects in these tasks, regardless of whether target trials are visual or audiovisual. In contrast, there are cross-modal links in the guidance of search for audiovisually defined targets, behavioural spatial cueing effects and/or N2pc components should be reduced or even completely eliminated in the audiovisual tasks, indicative of reduced ability of colour-cues to capture attention when they match only partly. Another important question was whether the dissociation between behavioural and ERP correlates of cue-triggered attentional capture observed previously during search for colour-size targets (Kiss et al., 2013) would also be found for audiovisual search. Lastly, a different pattern of spatial cueing effects and N2pc component results was expected across two audiovisual search tasks.

Method

Participants

Twelve right-handed paid volunteers with normal or corrected vision (mean age 25.8 years, age range 21–40 years, 5 females) took part. Informed consent was obtained from all participants prior to the start of the experiment.

Stimuli and apparatus

Visual stimuli were presented to participants at a distance of 100 cm on a 22" LCD monitor (Samsung wide SyncMaster 2233; 100 Hz refresh rate) against a black background. On each trial, a cue display (50 ms duration) was followed after a 150 ms interstimulus by a search array (50 ms duration). The intertrial interval was 1450 ms. Each cue and search array contained a circular array of six elements at an angular distance of 4.1° from a central fixation point (see Figure 5.X). Cue arrays contained six elements of four closely aligned dots (0.17° x 0.17°). One set of dots was a colour singleton that matched the target colour (blue or red, varied across subjects; CIE x/y chromaticity coordinates .161/.128 and .621/.128, respectively).

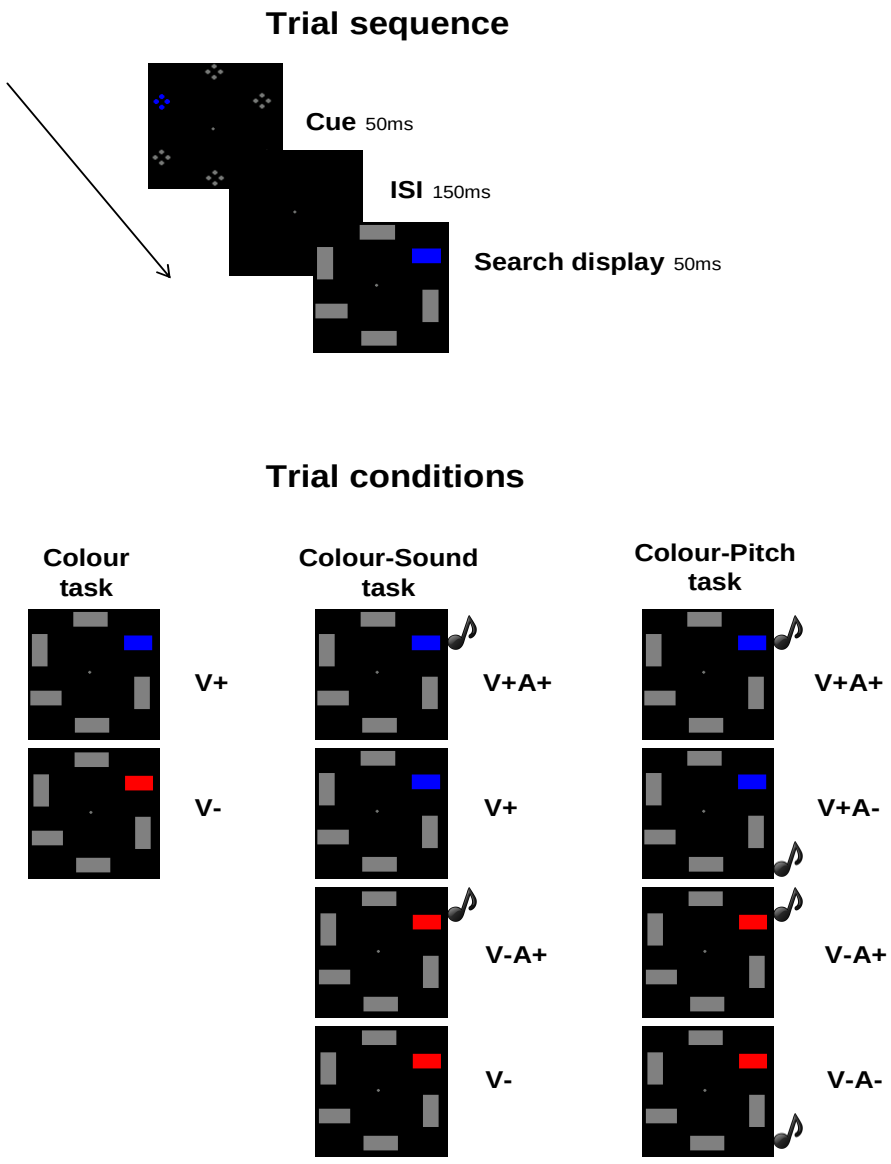

Figure 5.1. *Schematic illustration of the sequence of events on each trial (top panel), and the different search arrays presented in the Colour, Colour-Sound and Colour-Pitch tasks of Experiment 8, respectively (bottom panel). The example depicts a trial sequence and trial conditions across three tasks for participants who were searching for targets defined as blue bars (in the Colour task) and blue bars accompanied by high-pitch tones (in two audiovisual search tasks).*

The colour singleton was presented equiprobably and randomly at one of the four lateral locations, but never at the top or bottom position. The five remaining cue elements were uniformly grey (.308/.345). Search arrays contained six horizontal or vertical bars (1.1° x 0.3°) at the same positions as the preceding cue elements, with bar orientation chosen randomly for each position. One of these bars was always coloured (blue or red), the others were grey. Coloured bars appeared with equal probability at one of the four lateral locations. All grey, blue and red stimuli in the cue and search displays were equiluminant (~ 11 cd/m^2). In two of the three search tasks, search arrays could be accompanied by synchronous auditory stimuli. These were pure sine-wave tones (50 ms duration; 65 dB SPL; high-pitch: 2000 Hz; low-pitch: 300 Hz) that were presented concurrently with search array onset from a loudspeaker located centrally behind the monitor.

Procedure

Each participant completed three search task conditions. Participants were instructed to search for a target bar that was defined by a specific colour or by a specific colour-sound combination, to respond to the orientation of this bar by pressing one of two vertically aligned response keys, and to refrain from responding on nontarget trials. In the unimodal 'Colour' task, participants had to respond to a bar of one pre-specified colour (e.g., blue) and ignore bars of another, nontarget colour (e.g., red). The assignment of target and nontarget colours was counterbalanced across participants, and remained constant across all three tasks for each participant. Both trial types were presented with equal probability and in random order within each block. In the audiovisual 'Colour-Pitch' participants were instructed to respond to the orientation of target-colour bars only on trials where they were accompanied by a specific pitch (e.g., a blue bar accompanied by a high-pitch tone). These target trials (V+A+; e.g., a blue bar accompanied by high-pitch tone) made up 50% of all trials in each block. In this audiovisual task, all search arrays were accompanied by synchronous tones and nontarget trials were defined as trials where either the colour of the singleton bar (V-A+ trials; e.g. a red bar with a high-pitch tone), the pitch of the tone (V+A- trials; e.g., a blue bar with a low-pitch tone), or both did not match the target-defining feature (V-A- trials; e.g., a red bar with a low-pitch tone). The assignment of target and nontarget tone frequencies was counterbalanced across participants. In the 'Colour-Sound' audiovisual task, target trials were identical to the V+A+ trials in the Colour-Pitch task (also 50% of all trials). Nontarget trials in this audiovisual task were defined as trials where a target-colour or nontarget-colour bar appeared without tone (V+ and V- trials; e.g., a unimodal blue or red bar), or trials

where the sound with target-matching pitch accompanied a nontarget colour bar (V-A+ trials; e.g., a red bar with high-pitch tone). The three nontarget trial types were equiprobable in both audiovisual tasks.

Four successive blocks were run for each task, which were preceded by two training blocks. The order of the three search tasks was counterbalanced across participants. Each block included 96 trials, and 48 of these were target trials. Vertical and horizontal target bars were mapped onto the top and bottom key, respectively. The assignment of the left or right hand to the top or bottom response key was also counterbalanced across participants.

EEG recording and data analysis

EEG was DC-recorded from 23 scalp electrodes mounted in an elastic cap at standard positions of the extended 10-20 system at sites, Fpz, Fz, F3, F4, F7, F8, FC5 FC6, Cz, C3, C4, T7, T8, CP5, CP6, Pz, P3, P4, P7, P8, PO7, PO8 and Oz (500Hz sampling rate; 40Hz low-pass Butterworth filter). All scalp electrodes were online referenced to the left earlobe and re-referenced offline to the average of both earlobes. Impedances were kept below 5 kΩ. Horizontal eye movements (HEOG) were measured from two electrodes placed at the outer canthi of the eyes. Only trials with correct responses to the target were analyzed. Trials with saccades (voltage exceeding ± 30 µV in the HEOG channel), eyeblinks (exceeding ± 60 µV at Fpz) or muscle artefacts (exceeding ± 80 µV at any other electrode) were excluded from the analyses, as were trials with incorrect responses, missed targets, or False Alarms.

The EEG in response to the cue stimuli was epoched and averaged for the 500 ms interval after cue onset, relative to a 100 ms pre-cue baseline. Averages were computed for trials with colour singleton cues in the left and right hemifield, separately for all three tasks. N2pc amplitudes were quantified on the basis of mean amplitudes obtained between 170 ms and 270 ms after cue onset at lateral posterior electrodes PO7 and PO8. Onset latencies of cue-elicited N2pc components were compared across the three search tasks, using the jack-knife method described by Miller, Patterson, and Ulrich (1998) on the basis of difference waveforms obtained by subtracting ipsilateral from contralateral ERPs. The procedure estimates the onset latencies from grand averages for subsamples of participants successively excluding one participant from the original sample. N2pc onset latency was defined relative to an absolute amplitude criterion of -1 µV (see Eimer et al., 2011, for similar procedures), and *t* values were corrected according to the formula given by Miller and colleagues (1998). In all analyses, Greenhouse-Geisser corrections for violated sphericity assumptions were applied where appropriate.

Results

Behavioural performance

Trials with anticipatory and exceedingly slow responses were excluded from analyses, resulting in a loss of less than 1% all data. Figure 5.2 shows RTs for correct responses to targets at cued and uncued locations, separately for the three search tasks.

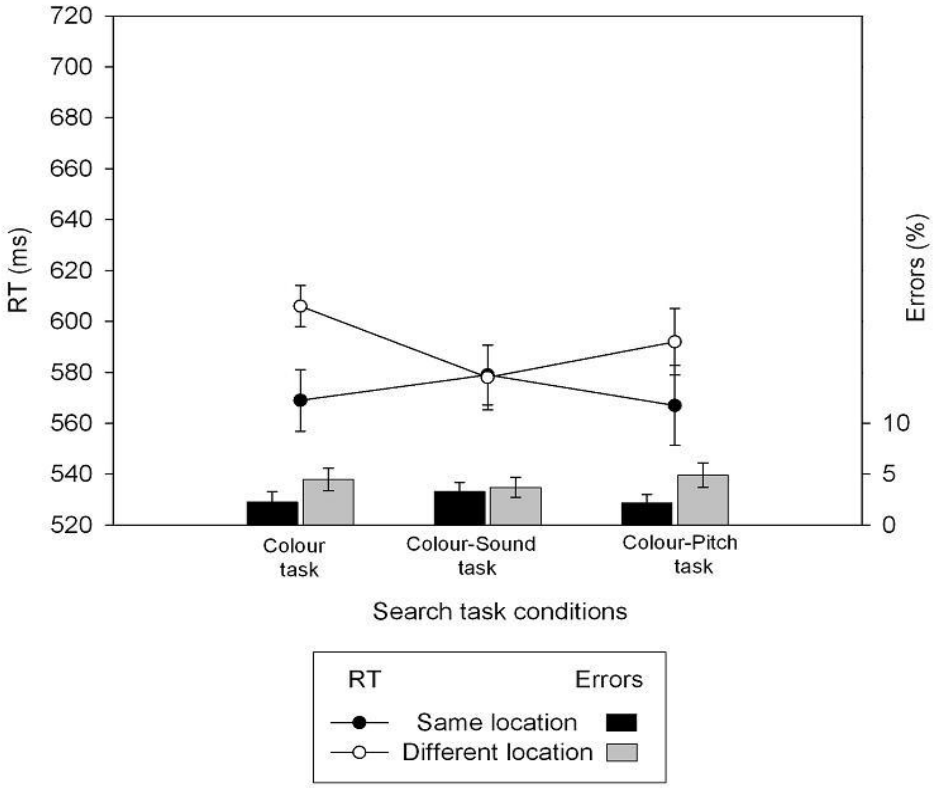

Figure 5.2. *Mean RTs (line graphs) and error rates (bar graphs) in Experiment 8 in response to targets at cued and uncued locations, shown separately for the Colour task, the Colour-Sound task, and the Colour-Pitch task. Error bars represent standard error of the mean.*

A repeated-measures ANOVA was conducted on the RT data for the factors spatial cueing (target at cued vs. uncued location) and task (Colour vs. Colour-Sound vs. Colour-Pitch). A lack of a main effect of task, $F < 1$, indicated that all tasks were performed with similar speed. A main effect of spatial cueing, $F(1,11) = 27.6$, $p < .001$, $\eta_p^2 = .72$, reflected faster RTs to targets at cued versus uncued locations. Most importantly, a two-way interaction between spatial cueing

and task was observed, $F(1.35,14.83) = 9.42$, $p < .01$, $\eta_p^2 = .46$, demonstrating that behavioural cueing effects triggered by the same target-colour cues differed between the unimodal visual and the two audiovisual search tasks. In the Colour task, a significant spatial cueing effect of 37 ms was observed, $F(1,11) = 46.5$, $p < .001$, $\eta_p^2 = .81$. In the Colour-Sound task, this effect was completely eliminated (-1 ms; $F < 1$). Planned comparisons confirmed that RT cueing effects in the Colour task were indeed reliably reduced in the Colour-Sound task, $F(1,11) = 13.23$, $p < .01$, $\eta_p^2 = .55$. In the Colour-Pitch task, reliable spatial cueing effects of 25 ms were observed, $F(1,11) = 23.6$, $p < .001$, $\eta_p^2 = .68$. Pair-wise comparisons demonstrated that cueing effects also in the Colour-Pitch task were reliably reduced when compared with the Colour task, $F(1,11) = 5.79$, $p < .05$, $\eta_p^2 = .35$.

A lack of a main effect of task, $F < 1$, suggested that errors occurred with a similar frequency in all three tasks. Response errors were more frequent to targets at uncued locations relative to cued targets (4.4% versus 2.6%; $F(1,11) = 6.26$, $p < .05$, $\eta_p^2 = .46$), but they were not further modulated by task, $F < 1$. Participants missed less than 1% of all targets on Go trials. False Alarms occurred on 1.2% of all Nogo trials, and False Alarm rates differed between the three search tasks, $F(2,22) = 6.23$, $p < .01$, $\eta_p^2 = .36$. False Alarms were virtually absent in the unimodal visual task (0.03%) and were relatively more frequent in the two audiovisual tasks (1.4% and 2.1% in Colour-Sound and Colour-Pitch task, respectively). In these audiovisual tasks, False Alarms were exclusively observed on trials where target-colour bars were presented (V+ trials or V+A- trials, respectively).

N2pc results

Figure 5.3 (top panels) shows ERPs triggered in response to cue arrays in the 350 ms interval after cue onset at electrodes PO7/8 contralateral and ipsilateral to the side of the colour singleton cue, separately for the Colour task, the Colour-Sound task and the Colour-Pitch task.

The N2pc component was triggered in response to colour singleton cues in all three search tasks. However, identical target-colour cues showed N2pc amplitude and onset latency differences across the three search tasks, which is clearly visible in the difference waveforms shown in Figure 5.3 (bottom panel) that were obtained by subtracting ipsilateral from contralateral ERPs. The N2pc appears to be larger and emerge slightly earlier in the unimodal colour task relative to the two audiovisual tasks.

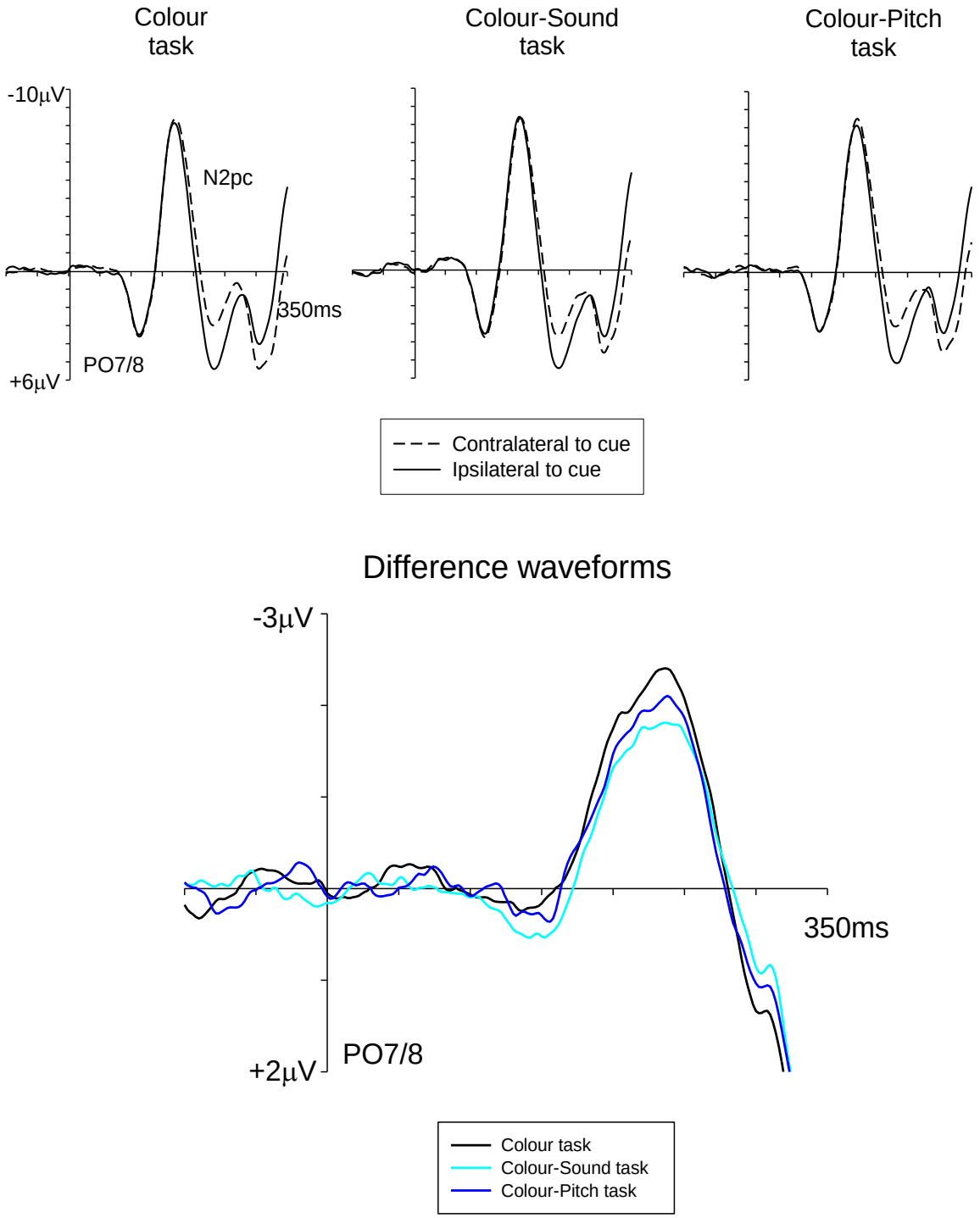

Figure 5.3. *Top panel: Grand-average ERPs measured in Experiment 8 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target-colour singleton cue, separately for the Colour task, the Colour-Sound task, and the Colour-Pitch task. Bottom panel: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the three search tasks.*

These differences were evaluated with a repeated-measures ANOVA for the factors contralaterality (electrode ipsilateral versus contralateral to the colour singleton cue) and task (Colour, Colour-Sound, Colour-Pitch). A main effect of contralaterality on N2pc mean amplitudes, $F(1,11) = 21.72$, $p < .001$, $\eta_p^2 = .66$, suggested the presence of N2pc components in response to cue arrays in all three search tasks. This was confirmed by pair-wise comparisons, with the smallest F value, $F(1,11) = 17.63$, $p < .001$, $\eta_p^2 = .62$, observed for a cue-induced N2pc component in the Colour-Pitch task. Crucially, there was a two-way interaction between contralaterality and task, $F(2,22) = 3.76$, $p < .05$, $\eta_p^2 = .26$, demonstrating that N2pc amplitudes differed across the unimodal and two audiovisual tasks. Planned comparisons confirmed that the N2pc elicited by cue arrays in the Colour-Sound task was reliably smaller than the N2pc measured in the unimodal visual task, $F(1,11) = 8.21$, $p < .01$, $\eta_p^2 = .43$. There was also a tendency for a reduction in the N2pc amplitude between the Colour task and the Colour-Pitch task, but this reduction failed to reach significance, $F(1,11) = 1.65$, $p = .11$. The N2pc component emerged significantly earlier in the unimodal Colour task than in the Colour-Sound task (185 ms versus 194 ms; $t_c(11) = 2.69$, $p < 0.05$). The N2pc onset latency difference between the Colour task and the Colour-Pitch task (185 ms versus 191 ms) failed to reach significance, $t_c(11) = 1.38$, $p = .098$.

Discussion

The aim of Experiment 8 was to investigate whether in tasks where targets are defined by features from two different modalities, attentional object selection is controlled by fully integrated bimodal object templates or separate within-modality attentional templates. For this purpose, the ability of a unimodal colour-matching singleton cue to attract attention in a task-set contingent fashion was compared across tasks in which the targets were defined by a single visual feature (i.e., colour) or by a conjunction of visual and auditory features (i.e., colour and pitch). Attentional capture elicited by unimodal colour-matching singleton cues, as indexed by spatial cueing effects and the N2pc components, was strongly reduced or eliminated altogether when targets were audiovisual. This remarkable pattern of results provides the first evidence that in contexts where targets are defined by a combination of features from more than one modality, task-relevant features are represented as fully integrated bimodal object templates that reduce the ability to capture attention of irrelevant distractors when they match only one of the target-defining features.

As expected, target-colour singleton cues captured attention in the unimodal Colour task, as reflected by behavioural spatial cueing effects and N2pc components triggered by unimodal cues. These results are in line with previous behavioural and electrophysiological evidence for task-set contingent attentional capture (e.g., Eimer & Kiss, 2008; Folk et al., 1992). In spite of the fact that the colour singleton cues were physically identical across all three tasks, behavioural attentional capture effects were substantially reduced in the two audiovisual relative to the unimodal Colour task. In the Colour-Sound task, RT spatial cueing effects were completely eliminated. In the Colour-Pitch task, they were significantly reduced relative to the unimodal task. Along similar lines, the N2pc triggered by colour singleton cues was significantly reduced in amplitude and delayed in onset in the Colour-Sound task when compared to the Colour task. Trends towards a reduction and delay of the N2pc component were also observed for the Colour-Pitch relative to the unimodal Colour task, although they did not reach statistical significance. If selection was controlled by separate, modality-specific attentional templates representing task-relevant visual and auditory features, behavioural and electrophysiological correlates of attentional capture by colour-singleton cues in these tasks should have been identical to the effects observed in the unimodal Colour task, as unimodal singleton cues always matched the target-defining colour. The observation that behavioural spatial cueing effects were reduced in the Colour-Pitch task and entirely absent in the Colour-Sound task, and the fact that the N2pc to colour singleton cues was attenuated and delayed in the Colour-Sound task strongly suggest that the ability of unimodal cues to capture attention was reduced during audiovisual search. These observations point to an important role for integrated bimodal object templates in the control of search for audiovisually defined targets. While consistent with the research indicating enhanced selection of bimodal object matching both features of naturalistic multimodal targets (Iordanesco et al., 2008, 2010), these findings are the first to demonstrate the flexibility of top-down mechanisms across modalities in the domain of feature-based attentional control (cf., Eimer et al., 2002). This issue will be addressed in more detail in the General Discussion.

Similar to the previous study on purely visual integrated object templates (Kiss et al., 2013), Experiment 8 revealed a dissociation between behavioural and electrophysiological markers of attentional capture in the Colour-Sound task. Behavioural spatial cueing effects were completely absent in this task, suggesting that target-colour singleton cues failed to capture attention. However, while the N2pc component triggered by colour-matching was attenuated when compared to the unimodal colour search task, it remained reliable, indicating that target-colour cues retained some of their ability to attract attention. This difference between electrophysiological and behavioural measures suggests that they reflect different aspects of task-

set contingent attentional capture also in audiovisual search tasks: The N2pc results support the explanation that argues that the initial selection stage is controlled by separate features, which enabled a unimodal distractor possessing one of the target-defining features to capture attention in a task-set contingent fashion. In contrast, the second stage of selection, indexed by behavioural spatial cueing effects, is under control of fully integrated bimodal target templates that triggered rapid disengagement from the *nontarget*-object location, as reflected by an absence of behavioural cueing effects in the Colour-Sound task. Thus, the current findings support the validity of the two-stage model of selection in task-set contingent capture proposed recently by Kiss et al. (2013), by demonstrating that it explains search behaviour also in environments in which targets are defined as conjunctions of features from more than one sensory modality.

While behavioural and electrophysiological markers of attentional capture were both reliably reduced in the audiovisual Colour-Sound task relative to the unimodal Colour task, the attenuation of cue-induced capture effects was less pronounced in the Colour-Pitch task, where the N2pc reduction only approached statistical significance. Why was attentional capture triggered by visual singleton cues more strongly reduced in the Colour-Sound task? The fact that errors and RTs were comparable across two audiovisual search tasks suggests that task difficulty is unlikely to explain this pattern of results. A more likely interpretation is enhanced top-down suppression of attentional capture by target-colour singleton cues in the Colour-Sound task. In this task, the cues were perceptually similar to the nontarget trials on which the search arrays with target-colour singleton bars were presented without a synchronous sound (V+ trials; see Figure 5.1, bottom panel). In contrast, in the Colour-Pitch task, where all search arrays were accompanied by tones, the features of the cue arrays (visual colour singletons without synchronous tones) did not correspond to any nontarget trial type, which may have resulted in a weaker inhibition of attentional capture. Thus, similarity to nontarget stimuli, rather than generic task difficulty, seemed to underlie the divergent pattern of task-set contingent capture by partly matching distractors across two audiovisual search tasks. This issue will be discussed in more detail in the General Discussion.

The explanation presented above may account for the differences between the two audiovisual tasks in Experiment 8. However, the fact remains that electrophysiological attentional capture effects did not differ reliably between the unimodal Colour task and the audiovisual Colour-Pitch task. This may cast doubt as to whether integrated bimodal attentional templates play a central role in the guidance of search for audiovisual targets. In Experiment 9, participants were given a stronger incentive to treat the auditory target-defining attribute in the Colour-Pitch task as more relevant to the task-at-hand.

Experiment 9. The role of task-dependent relevance of the target-defining features in task-set contingent capture in audiovisual search tasks

Introduction

The results from Experiment 8 demonstrated that the selection of visual stimuli in space can be guided by fully integrated bimodal object templates. This suggests the presence of a top-down mechanism, by which attentional weights applied to inputs from channels coding specific visual features projected onto the saliency map responsible for allocation of attention in space (Wolfe, 1994, 2007) can be flexibly adjusted to reduce the ability to capture attention of objects matching only one of the features of an audiovisually defined target. While this reduction was present in the Colour-Pitch task, as indexed by attenuated and delayed N2pc components, it was not reliable in the Colour-Sound task, which might weaken the conclusion that bimodal search templates play a central role in the control of visuo-spatial attention.

In contexts of search for targets defined by conjunctions of features from different modalities, channels coding the two target-defining features might receive different attentional weights, as determined by the specific demands of task. This could explain the stronger ability of task-irrelevant objects to capture attention in the Colour-Pitch task, where they were not explicitly defined as distractors, and thus may not have been subjected to enhanced top-down inhibition. In line with the proposed flexibility of top-down control mechanisms, Bacon and Egeth (1997) demonstrated that in order to improve performance in searching for visual feature conjunctions, search can be flexibly restricted in a task-dependent fashion to one target-defining feature. In one of their studies, Bacon and Egeth (1997; Experiment 2) instructed participants to restrict search for colour-orientation targets to one of the two target-defining features (e.g., 'restrict your search to the *red* elements'), and highlighted that, because there were fewer distractors that matched this feature, the presence of a particular feature would be more strongly indicative of a target. Results showed that search times were shorter on trials that were consistent relative to inconsistent with the advised search strategy, suggesting flexibility in the adjustment of weights to inputs from channels coding different target-defining features of a multi-feature search template.

Experiment 9 was designed to investigate whether encouraging participants to regard the auditory target-defining feature of the audiovisual target in the Colour-Pitch task as more task-

relevant would result in more reliable effects of integrated bimodal templates, as indexed by reduced spatial cueing effects as well as attenuated and delayed N2pc components. For this purpose, the proportion of nontarget trial types was altered in the current experiment. Relative to Experiment 8, the number of nontarget trials where a target-pitch tone was presented simultaneously with a nontarget-colour bar (V-A+ trials) was reduced from 16 to 4, while the number of trials with target-colour bars accompanied by nontarget-pitch tones (V+A- trials) was increased from 16 to 28. As a result, the presence of the target-pitch sound was now much more strongly associated with the target status of a given trial. This manipulation should result in a stronger role for the auditory features in the control of object selection by integrated bimodal attentional templates in the Colour-Pitch task, and thus in a reliable reduction of both behavioural and electrophysiological markers of attentional capture relative to the unimodal Colour task.

Method

Participants

Thirteen paid volunteers took part in the study. Data from one participant was not included in the analyses due to excessive activity in the alpha band. The remaining twelve participants (mean age 28.5 years, age range 22–38 years; 1 left-handed; 5 males) had normal or corrected-to-normal vision. All gave informed consent to participate in the study.

Stimuli, procedure, and design

Experimental procedures were identical to Experiment 8 with a few notable exceptions. First, only the Colour task and the Colour-Pitch task were conducted. Second, in the Colour-Pitch task, the proportion of trials with distractors with target pitch (V-A+), as well as with the distractors with target colour (V+A-), was changed in a way to make the presentation of target-pitch sound (A+) more indicative of presence of the audiovisual target. Namely, in Experiment 8, each block of 96 trials included 16 trials with nontarget-colour/target-pitch (V-A+) distractors and 16 trials target-colour/nontarget-pitch (V+A-) distractors. Hence, with a bimodal target (V+A+) presented on 48 trials per block, the probability that the presentation of the target feature (either colour or pitch) indicated the presence of the audiovisual target on any given trial was 3 to 1. In Experiment 9, there were only 4 trials with nontarget-colour/target-pitch (V-A+) distractors, and 28 trials

target-colour/nontarget-pitch (V+A-) distractors. As a result of this manipulation, now the presentation of sound with the target-pitch (A+) was associated with the presence of the audiovisual target on 48 out of 52 trials per block, thus increasing the probability of the target presence being indicated by the presentation of target-pitch sound to 12 to 1. In contrast, the strength of the association between the presence of the target colour (V+) and the target status of a given trial was now reduced: There were 76 trials where a target-colour bar was present, but only 48 of these required a response, thus reducing the probability of target presence being indicated by presence of the target colour to approximately 3 to 2. The number of V+A+ and V-A- trials per block (48 versus 16) remained unchanged. Participants were informed that distractors sharing the target-pitch (V-A+) were quite rare (but their exact number was not revealed), and that search would be easier if they try to focus in their search for bimodal target on stimuli with the target-pitch sound (cf., Bacon & Egeth, 1997). To prevent participants from adopting a unimodal auditory task set, participants were explicitly instructed not to make a response on the rare trials where nontarget-colour/target-pitch (V-A+) distractors were presented.

EEG recording and data analysis

EEG recording and analysis procedures were identical to the ones employed in Experiment 8, except that the search task was now a two-level factor (i.e., Colour task vs. Colour-Pitch task).

Results

Behavioural performance

Trials with anticipatory and exceedingly slow responses were excluded, resulting in a loss of less than 1% of all trials. Figure 5.4 depicts RTs for correct responses and error rates for targets presented at cued and uncued locations, separately for the two search tasks. A two-way ANOVA with spatial cueing and search task as within-subject factors revealed a tendency for faster responses in the Colour compared to Colour-Pitch task (597 ms vs. 610 ms), $F(1,11) = 4.25$, $p = .064$, $\eta_p^2 = .28$. A main effect of spatial cueing, $F(1,11) = 64.51$, $p < .001$, $\eta_p^2 = .85$, suggested reliable spatial cueing effects in both search tasks. This was confirmed by pair-wise comparisons, which revealed significant cueing effects of 42 ms in the Colour task ($F(1,11) = 95.1$, $p < .001$, $\eta_p^2 = .9$) and of 30 ms ($F(1,11) = 23.74$, $p < .001$, $\eta_p^2 = .68$) in the Colour-Pitch task. Importantly, as in Experiment 8 cueing effects elicited by target-colour cues were reliably reduced when the Colour and the Colour-Pitch search task were compared (see Figure 5.4, line graphs), as

evidenced by a two-way interaction between search task and spatial cueing, $F(1,11) = 4.98$, $p < .05$, $\eta_p^2 = .31$.

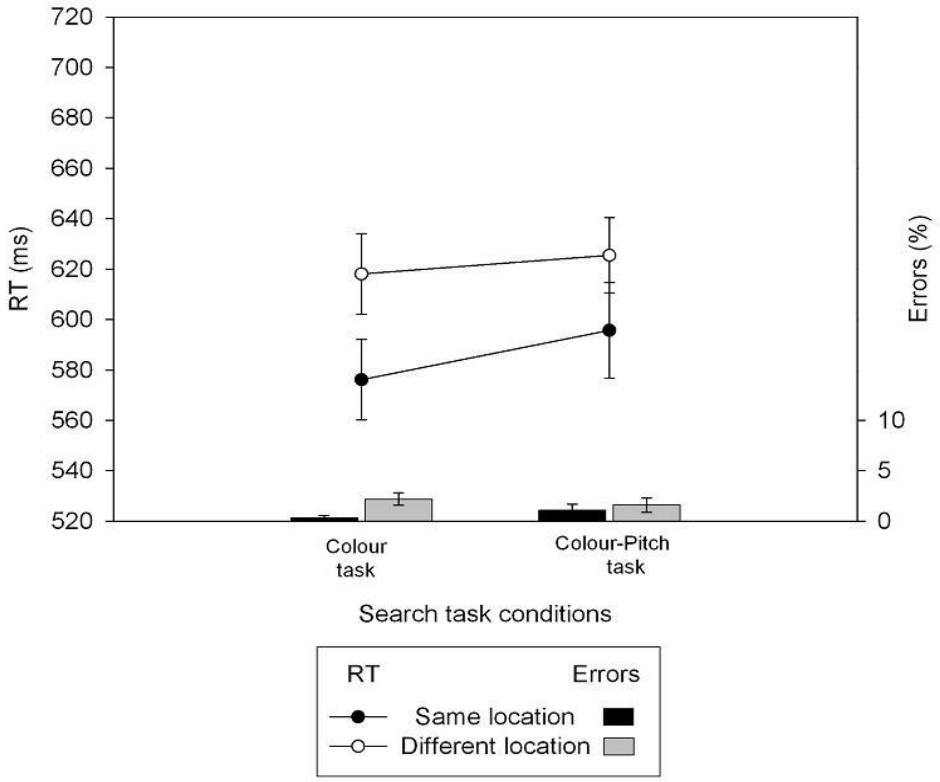

Figure 5.4. *Mean RTs (line graphs) and error rates (bar graphs) in Experiment 9 in response to targets at cued and uncued locations, shown separately for the Colour task and the Colour-Pitch task. Error bars represent standard error of the mean.*

As shown by Figure 5.4 (bar graphs), erroneous responses were more frequent when targets were presented at uncued relative to cued locations (1.9% vs. 0.7%), $F(1,11) = 4.89$, $p < .05$, $\eta_p^2 = .31$. There was no main effect of task on error rates, $F < 1$, and no interaction between task and spatial cueing, $F(1,11) = 2.5$, $p = .14$. Participants missed less than 1% of all targets on Go trials and failed to respond on less than 1% of all target trials. False Alarms occurred on average on 0.3% of all Nogo trials and were not modulated by task, spatial cueing or an interaction of the two factors (the smallest $p > .21$).

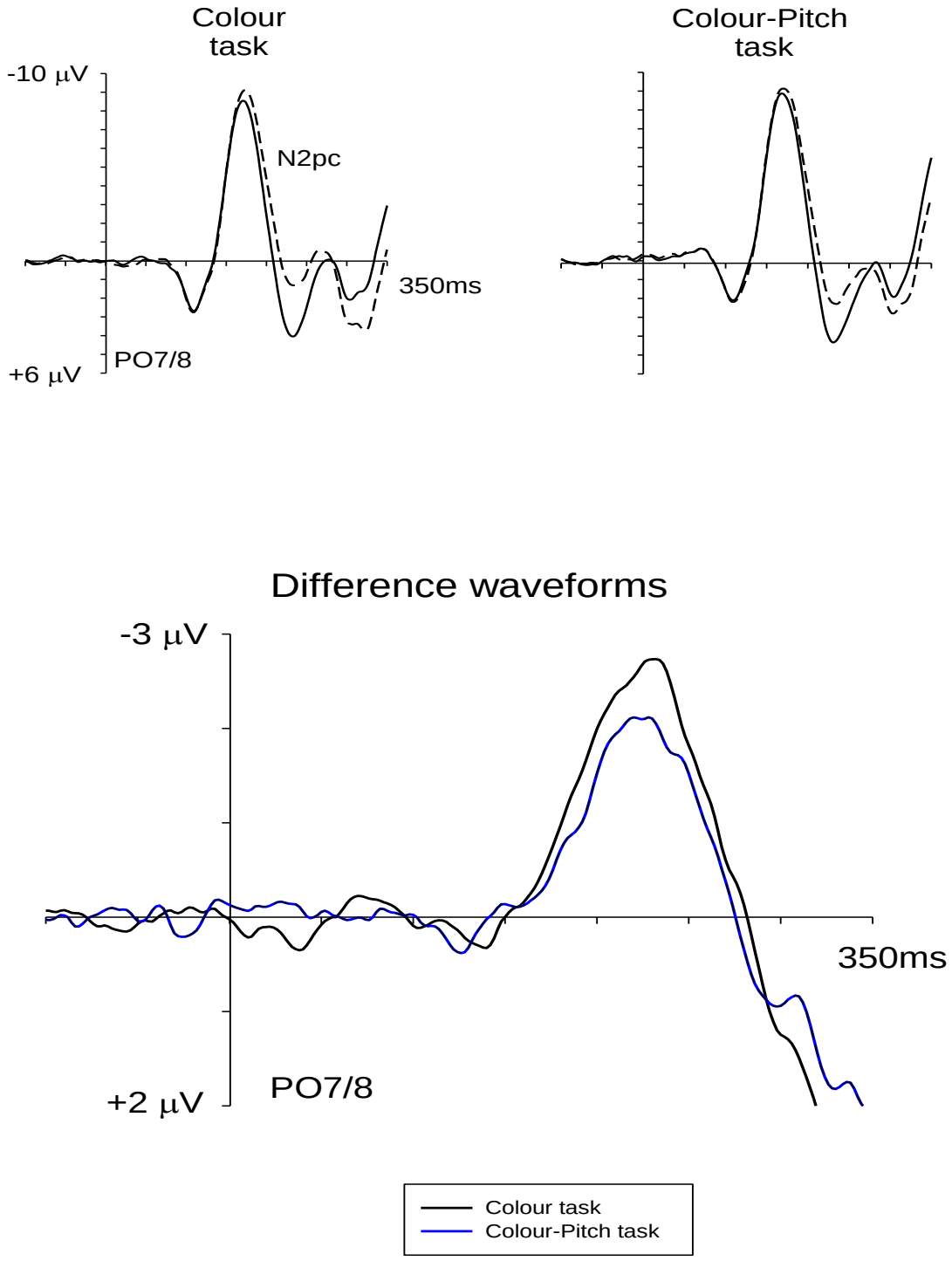

Figure 5.5. *Top panel: Grand-average ERPs measured in Experiment 9 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target-colour singleton cue, separately for the Colour task and the Colour-Pitch task. Bottom panel: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the two search tasks.*

N2pc results

Figure 5.5 (top panels) shows grand-averaged ERPs triggered in response to cue arrays in the 350 ms interval after cue onset at electrodes PO7/8 contralateral and ipsilateral to the side of the colour singleton cue, separately for the Colour and the Colour-Pitch task. As in Experiment 8, N2pc components were triggered in response to the colour singleton cues in both search tasks, as shown by the differences waveforms obtained by subtracting ipsilateral from contralateral ERPs (Figure 8, bottom panel). Importantly, in contrast to Experiment 8, N2pc amplitudes and onset latencies in the Colour-Pitch task were now markedly different from the Colour task. Mean ERP amplitudes recorded in the 170–270 ms time window after onset of the colour singleton cues were analysed with a repeated-measures ANOVA for factors contralaterality and task. A main effect of contralaterality, $F(1,11) = 43.37$, $p < .001$, $\eta_p^2 = .8$, suggested that reliable N2pc components were elicited in both search tasks. This was confirmed by pair-wise comparisons, with a significant N2pc component found in the Colour task, $F(1,11) = 44.41$, $p < .001$, $\eta_p^2 = .8$, as well as in the Colour-Pitch task, $F(1,11) = 35.87$, $p < .001$, $\eta_p^2 = .77$. Crucially, a two-way interaction between contralaterality and task was observed, $F(1,11) = 9.68$, $p < .01$, $\eta_p^2 = .47$, which provided evidence that in Experiment 9 N2pc amplitudes in response to colour singleton cues were reliably attenuated in the Colour-Pitch task relative to the Colour task. This is clearly visible in the difference waveforms in Figure 5.5 (bottom panel). The N2pc onset latency in the Colour-Pitch task (191 ms) was now significantly delayed compared to the Colour task (181 ms), $t_c(11) = 1.84$, $p < .05$.

Discussion

In Experiment 9, the probability of different nontarget trial types was altered in a way that increased the task-relevance of target-defining auditory features in the Colour-Pitch task. As a consequence of this manipulation the presence of the target-defining pitch was more strongly associated with the target status of a given trial in Experiment 9 compared to Experiment 8. A reliable reduction of attentional capture by target-colour cues in this audiovisual task context, as compared to the unimodal visual task, was now observed with both behavioural and ERP measures. RT spatial cueing effects were significantly smaller in the Colour-Pitch task than in the unimodal Colour task. In contrast to Experiment 8, where these reductions were only at a trend level, the N2pc to target-colour singleton cues in this audiovisual task was now reliably attenuated and delayed in onset as compared to the unimodal task.

The observed reductions further strengthen the hypothesis that spatial selection can be effectively controlled by integrated object templates in contexts where search targets are defined by features from different modalities. Both behavioural and ERP results demonstrate that attentional capture by colour-matching unimodal singleton cues can be reliably attenuated during audiovisual search even in contexts where such distractors are not explicitly defined as distractors. These findings provide evidence that weights assigned to inputs provided by channels coding different target features onto the central salience map responsible for the allocation of visual attention in space can be readily adjusted via a top-down knowledge-based mechanism also in audiovisual search tasks (cf., Bacon & Egeth, 1997). However, although both behavioural and ERP measures of attentional capture now showed reliable reductions in the Colour-Pitch task, they were still significant, which suggests that unimodal target-colour cues retained some of their ability to attract attention in this audiovisual task context. Interpreted within the dual-stage model proposed by Kiss et al. (2013), these results indicate that top-down control mechanism based on the knowledge about audiovisual distractor frequency does not prevent distractors from triggering initial attentional capture (due to matching one of the target-defining features) or holding attentional focus in its location. Care is warranted when interpreting the presence of reliable behavioural and ERP markers of attentional capture as evidence for limits in the effectiveness of audiovisual object templates in controlling attentional capture by visual feature-matching distractors. In the Kiss et al. study (2013; Experiment 1), where behavioural spatial cueing in response to partly-matching cues were eliminated in a visual feature-conjunction search task, these cues were perceptually similar to stimuli presented on nontarget trial types. Additional research is required to establish whether the similarity of partially task-set matching distractors to explicitly defined nontargets is a critical determinant for their residual ability to capture attention in unimodal and bimodal task contexts.

While both Experiment 8 and 9 demonstrated that attentional capture by colour-matching cues can be strongly attenuated in audiovisual search contexts, indicative of top-down control of selection by integrated bimodal search templates, the electrophysiological attentional capture effects triggered by such distractors were still reliable across the three audiovisual tasks employed. Experiment 10 was designed to investigate how relative salience of the partly-matching distractors modulates their ability to capture attention in contexts where targets are defined across modalities.

Experiment 10. The role of distractor salience in task-set contingent attentional capture in audiovisual search tasks

Introduction

Research conducted in the last two decades understanding of the role of bottom-up, salience-driven mechanism in the control of where the search targets were defined by a visually unique single feature has greatly enhanced the attentional selection (Eimer et al., 2009, 2010; Folk et al., 1992, 1998; Lamy et al., 2003, 2004; Theeuwes, 1991,1994). The ability of distractors that are salient (i.e., feature singletons) but do not match the target-defining feature to capture attention is either eliminated or strongly reduced by currently active top-down task sets (e.g., Folk et al., 1992). This is likely due to top-down inhibitory mechanisms that were revealed by studies using ERP techniques (Eimer et al., 2009; Hickey et al., 2008; Kiss et al., 2012; Sawaki & Luck, 2010). These studies have demonstrated that the role of visual salience in the control of visuo-spatial attention orienting is at best indirect, in that it may activate additional control mechanisms that prevent it from controlling the location of attentional focus. Additionally, visual salience was shown not to be necessary for task-set matching distractors to attract visuo-spatial attention to their location (Eimer et al., 2009; Lamy et al., 2004).

In this context, research investigating the mechanisms by which attention is controlled during search for audiovisually defined targets poses an important novel question in respect to the role of salience in task-set contingent capture in environments where targets are defined by conjunctions of features: Does within-modality salience result in a stronger or weaker reduction in the ability of partly-matching distractors to capture attention in conditions in which the targets is defined by a conjunction of features from different modalities? On a closer look, behavioural and ERP evidence from the visual domain suggests that whether salient but irrelevant distractors will trigger attention shifts, inhibition, or have no effect on the ongoing selection process can depend on their relative salience, with higher-salience irrelevant events, e.g., large or bright singletons, more reliably overriding attentional control settings (Eimer & Kiss, 2010; Folk & Remington, 1998; Lamy & Egeth, 2003; Yantis & Egeth, 1999). Importantly, while Lamy et al. (2004) demonstrated that in visual single-feature search, there are no differences in cueing effects triggered by higher-salience feature singletons and lower-salience heterogeneous target-matching

colour cues, Eimer et al. (2009) showed that comparable cueing effects triggered by such cues are accompanied by N2pc amplitudes that are reliably larger than ones triggered by heterogeneous cues, suggesting that former capture attention more strongly.

It is possible that in contexts, where targets are defined by a conjunction of visual and auditory features, the ability of lower-salience distractors that share one of the task-relevant features to capture attention in a task-set contingent fashion will be more strongly reduced when compared with higher-salience target-matching distractors. In other words, attentional capture effects triggered by lower-salience target-matching cues may be inhibited more effectively than capture effects by higher-salience cues. In order to assess whether lower relative salience of partly matching distractors results in stronger inhibition of attentional capture in an audiovisual task, Experiment 10 used procedures identical to the ones employed in Experiment 8, with the sole exception that the colour-matching cues in the cue display were now presented against a background of five differently coloured items. It was predicted that attentional capture triggered by such target-matching cues will be completely eliminated in audiovisual task contexts, as evidenced by elimination of spatial cueing effects as well as of the N2pc component.

Method

Participants

Thirteen participants took part in this experiment. One was excluded due to excessive eye movements. Twelve remaining participants (mean age 27.3 years, age range 21–37 years, 5 males) were all right-handed and had normal or corrected vision. All gave informed consent.

Stimuli, procedure, and design

The experimental procedures were identical to the ones employed in Experiment 8, except that target-colour cues were now presented against a heterogeneous distractor background (cf., Figure 5.1). Each of the five items in the cue display was now randomly assigned a different colour from a set of six task-irrelevant colours with different CIE chromaticity coordinates (purple .227/ .129; turquoise .252/.427; green .262/.558; pink .491/.289; orange .569/ .392; yellow .438/.452). All visual stimuli were equiluminant (~11 cd/m2).

EEG recording and data analysis

EEG recording and analysis procedures were identical to the ones employed in Experiment 8.

Results

Behavioural performance

Exclusion of trials with anticipatory and very slow responses led to a loss of less than 1% of all data. Figure 5.6 depicts RTs for correct responses and error rates for targets presented at cued and uncued locations, separately for the Colour task, the Colour-Sound task and the Colour-Pitch task.

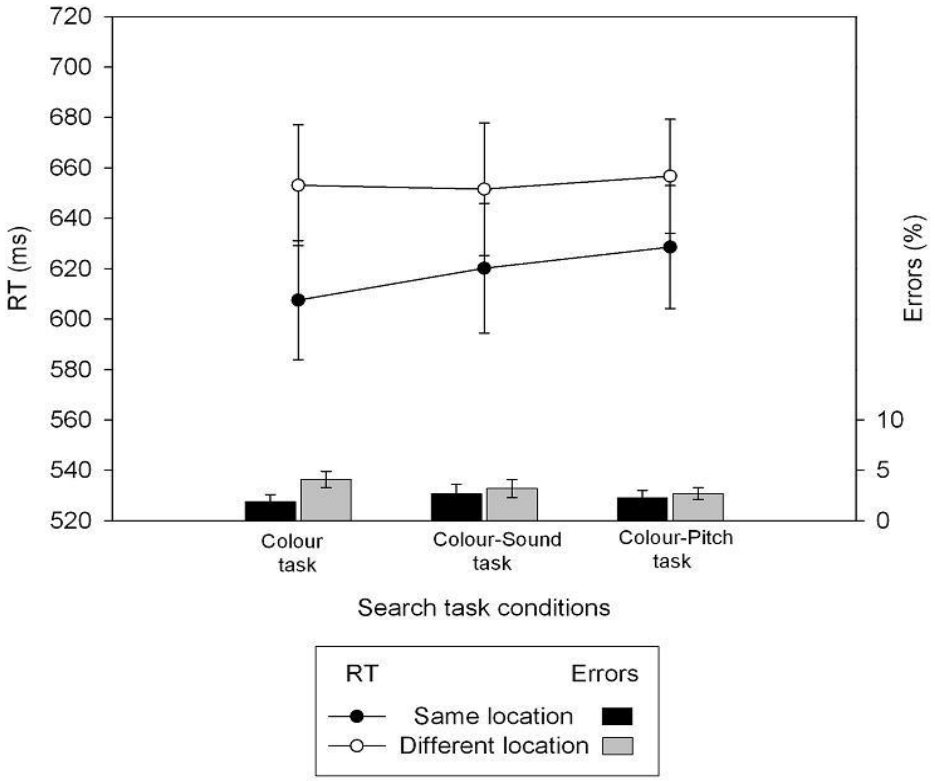

Figure 5.6. *Mean RTs (line graphs) and error rates (bar graphs) in Experiment 10 in response to targets at cued and uncued locations, shown separately for the Colour task, the Colour-Sound and the Colour-Pitch task. Error bars represent standard error of the mean.*

Participants performed all three search task with similar speed, as suggested by a lack of main effect of task on RTs, $F < 1$. A main effect of spatial cueing, $F(1,11) = 66.89$, $p < .001$, $\eta_p^2 = .86$, indicated that reliable cueing effects were elicited in response to target-colour cues in all three tasks. Planned comparisons confirmed this by showing significant cueing effects of 46 ms in the Colour task, $F(1,11) = 64.13$, $p < .001$, $\eta_p^2 = .85$, 31 ms in the Colour-Sound task, $F(1,11) = 28.64$, $p < .001$, $\eta_p^2 = .72$, and 28 ms in the Colour-Pitch task, $F(1,11) = 32.5$, $p < .001$, $\eta_p^2 = .75$. Importantly, also in the present experiment, where the colour cues were presented against a heterogeneous distractor background, the spatial cueing effects they triggered differed as a function of search task, evidenced by a two-way spatial cueing x task interaction, $F(2,22) = 4.89$, $p < .05$, $\eta_p^2 = .31$. As predicted, planned comparisons demonstrated that the reduction of cueing effects was observed when Colour task was compared both to the Colour-Sound task, $F(1,11) = 7.45$, $p < .01$, $\eta_p^2 = .4$, as well as the Colour-Pitch task, $F(1,11) = 6$, $p < .05$, $\eta_p^2 = .35$.

Erroneous responses were observed on average on 2.8% of all trials (see Figure 5.6). Error rates were not modulated by search task, spatial cueing or an interaction of these two factors (the smallest $p = .12$). Participants missed less than 1% of all targets on Go trials. False Alarms were found on average on 1.2% of all trials, and differed across three search tasks, $F(2,22) = 6.46$, $p < .01$, $\eta_p^2 = .37$. Pair-wise comparisons showed reliably fewer False Alarms when the Colour task (0.3%) was compared to the Colour-Pitch task (2.2%), $p < .05$, but not for the comparison between the Colour and the Colour-Sound task (1%), $p = .067$. The difference between the two audiovisual tasks also failed to reach significance, $p = .055$. False Alarms were not modulated by spatial cueing or by a spatial cueing x task interaction (smaller $p = .12$).

Combined analysis of Experiments 8 and 10

To further investigate the role of visual salience in the effects of bimodal task sets on behavioural capture effects, a mixed ANOVA was conducted on combined RTs data from Experiments 8 and 10 for the within-subjects factors search task and spatial cueing, and the between-subjects factor cue salience (high vs. low). A main effect of cue salience indicated that overall the search tasks were performed faster when the cue displays contained a singleton cue relative to heterogeneous cue (582 ms vs. 636 ms), $F(1,22) = 4.44$, $p < .05$, $\eta_p^2 = .17$. A two-way spatial cueing x cue salience interaction, $F(1,22) = 6.49$, $p < .05$, $\eta_p^2 = .23$, indicated that smaller cueing effects were observed overall in Experiment 8, where cues were colour singletons, relative to Experiment 10, where heterogeneous colour cues were employed. Critically, this effect was accompanied by a three-way task x spatial cueing x cue salience interaction, $F(2,44) = 4.15$, $p < .05$, $\eta_p^2 = .16$,

which suggested a role for cue salience as a factor modulating the task-set driven reduction of cueing effects triggered by unimodal colour-matching cues. Planned comparisons showed that the Colour versus the Colour-Pitch task reduction of cueing effects was not modulated by cue salience, $F < 1$. However, when the Colour and Colour-Sound tasks were compared, a stronger reduction of cueing effects was observed for singleton relative to heterogeneous target-colour cues, $F(1,22) = 4$, $p < .05$, $\eta_p^2 = .15$.

N2pc results

Figure 5.7 (top panels) depicts grand-averaged ERPs triggered in response to cue arrays in the 350 ms interval after cue onset at electrodes PO7/8 contralateral and ipsilateral to the side of the colour cue, separately for three search tasks. In contrast to Experiment 8, there were now no apparent N2pc amplitude and onset latency differences when the Colour-Sound and the Colour-Pitch task were compared to the Colour task. A repeated-measures ANOVA for the factors task and contralaterality that was carried out on N2pc mean amplitudes recorded in the 170–270 ms time window after onset of colour cues showed a main effect of contralaterality, $F(1,11) = 21.93$, $p < .001$, $\eta_p^2 = .67$. Planned comparisons revealed reliable cue-elicited N2pc components in all three search tasks *F's* > 16.4. Surprisingly, the N2pc amplitudes were comparable across the three search task, as suggested by a lack of interaction between contralaterality and task, $F < 1$.

Combined analysis of Experiments 8 and 10

The N2pc data recorded in the 170–270 ms post-cue time window from combined across Experiments 8 and 10 was analysed in a mixed ANOVA for the within-subjects factors task and contralaterality, and the between-subjects factor cue salience (high vs. low). A three-way task x contralaterality x cue salience interaction, $F(2,44) = 3.05$, $p < .05$, $\eta_p^2 = .12$, suggested that cue salience is an important factor modulating the reduction of capture effects as a function of search task, as indexed by attenuated N2pc component. Similarly to RTs results, planned comparisons demonstrated stronger N2pc amplitude reductions for singleton relative to heterogeneous target-colour cues when the Colour and Colour-Sound task were compared, $F(1,22) = 6.38$, $p < .01$, $\eta_p^2 = .23$, but no similar effect of cue salience for the Colour versus the Colour-Pitch task comparison, $F(1,22) = 1.23$, $p = .28$. No other interaction involving salience as a factor was significant, the smallest $p > .18$.

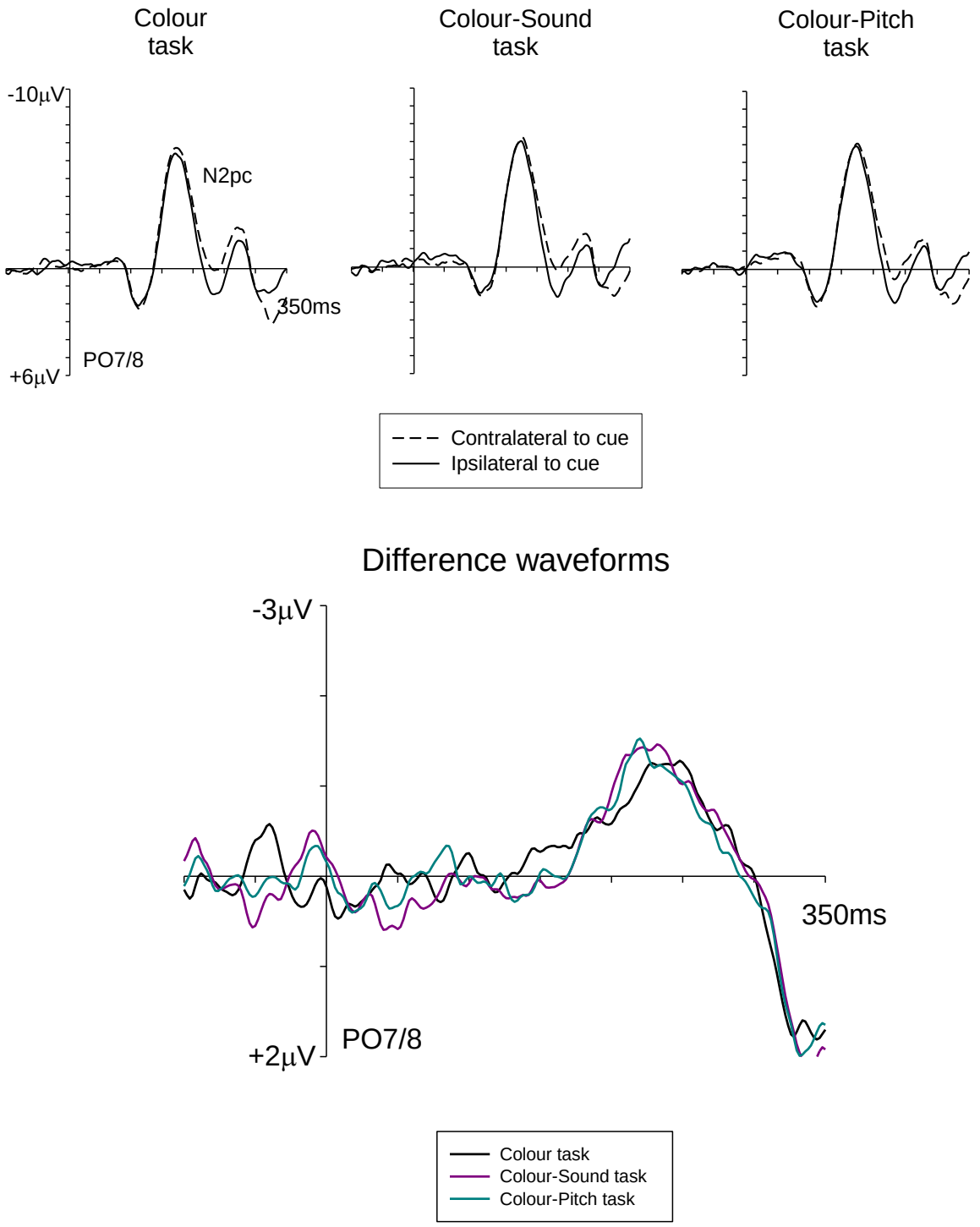

Figure 5.7. *Top panel: Grand-average ERPs measured in Experiment 10 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target-colour singleton cue, separately for the Colour task, the Colour-Sound task, and the Colour-Pitch task. Bottom panel: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the three search tasks.*

Discussion

The aim of Experiment 10 was to investigate the role of within-modality salience of colour-matching cues in contexts in which search can be guided by integrated audiovisual object templates. For this purpose, colour-matching cues were now presented against five differently coloured items in the cue array. Behavioural and electrophysiological markers of attentional capture showed disparate results. Behavioural spatial cueing effects triggered by heterogeneous colour-matching cues were reliably reduced in both Colour-Sound and Colour-Pitch task relative to the Colour task, but the mean amplitudes of the N2pc component triggered by unimodal colour-matching cues in the two audiovisual search tasks were comparable to the N2pc measured in the unimodal search task.

While the present divergent pattern of results is difficult to explain by an account in which the behavioural and ERP markers of attentional capture reflect the same selection process activated by irrelevant distractors sharing task-relevant features, these findings could be accommodated by the dual-stage selection model of task-set contingent capture (Kiss et al., 2013). Reduced behavioural cueing effects in the two audiovisual search tasks compared to the unimodal task are consistent with the control of this stage by integrated templates, where detection of a mismatch with the template results in the disengagement of the attentional focus from the distractor location. However, the reliability of behavioural capture effects in the Colour-Sound and Colour-Pitch tasks indicates that in both of these tasks in Experiment 10 the partly-matching distractors retained some of their ability to attract attention. The current N2pc results are also in line with the dual-stage selection account: The absence of N2pc amplitude differences between audiovisual versus visual search task is not surprising if the selection stage reflected by the N2pc component is controlled predominantly by input from independent and separate channels coding specific features.

If interpreted in line with the dual-stage selection model (Kiss et al., 2013), the pattern of behavioural and ERP results from Experiment 10 would suggest that audiovisual templates control selection of target-matching solely at the later stage associated with maintenance of attentional focus in distractor location, but it has no visible influence on the earlier stage during which selection occurs on the basis of presence of task-relevant features. However, these conclusions are inconsistent with the results of Experiment 8, where N2pc differences between both audiovisual search tasks and the unimodal task (although the N2pc attenuation and onset delay in the Colour-Pitch task failed to reach significance) suggested that integrated audiovisual templates control to some extent even the earlier stage of initial attentional capture. This *prima facie* confusing picture becomes much clearer when the behavioural and ERP indices of

attentional capture elicited by heterogeneous versus singleton colour-matching cues across Experiments 8 and 10 are compared. For both spatial cueing effects and the N2pc results, cue salience was an important factor in the reduction of capture effects as a function of task set. Critically, planned comparisons revealed that the modulatory role of salience was circumscribed to the Colour-Sound task, and it did not affect the pattern of results in the Colour-Pitch task across the two experiments.

In the Colour-Sound task, only singleton cues, but not heterogeneous cues, triggered reliably reduced spatial cueing effects and reduced N2pc components in the audiovisual compared to the unimodal visual task context. This pattern of results suggests that increased visual salience of partly task-matching cues results in a stronger reduction of their ability to capture attention during audiovisual search, but this effect is characteristic only of contexts in which such cues are perceptually similar to a nontarget stimulus. In contrast, in contexts in which target-matching unimodal cues are not perceptually similar to nontarget stimuli, lower-salience and higher-salience target-matching cues triggered comparable attentional capture effects across unimodal and audiovisual contexts, as indexed by both behavioural and ERP measures. This indicates that the mismatch between a partly-matching object and the audiovisual object template (cf., Duncan & Humphreys, 1989) might be easier to detect in circumstances when this object is salient. These findings provide the first evidence that at the stage at which the selection in task-set contingent capture is controlled by input from separate feature channels, salient objects will have a reduced ability to capture attention in bimodal task contexts.

To sum up, the results of Experiment 10 highlighted an important role of within-modal salience in task-set contingent capture in audiovisual search task contexts. Perceptual salience of partly matching irrelevant objects facilitates suppression of input from the channels coding the feature they match, and support the disengagement of attentional focus from the location of such partly matching distractor objects. The critical novel finding from Experiment 10 is that the role of relative visual salience in facilitating attentional control in multi-feature search contexts might be contingent on the activation of top-down inhibition mechanisms towards unimodal target-matching irrelevant objects (cf., Experiment 11). Further research is required to determine whether this hierarchy between salience-driven and goal-based factors applies also to task-set contingent capture in search contexts guided by object templates defined by features coded in the same modality.

Experiment 11. Top-down guidance of search for size-pitch feature conjunctions by audiovisual object templates

Introduction

Experiments 8 and 9 demonstrated that task-set contingent attentional capture by colour singleton cues is reduced during search for audiovisually as compared to purely visually defined targets, thereby providing strong evidence for top-down control of audiovisual search by integrated bimodal attentional templates. Behavioural and ERP capture effects triggered by target-matching cues were reliably reduced in task contexts where the cue was perceptually similar to nontarget stimuli in the search array (Experiment 8, Colour-Sound task) or when the auditory target-defining feature strongly indicated the presence of audiovisual targets (Experiment 9). Interestingly, in the study of Kiss et al. (2013; Experiment 1), the N2pc components elicited by partially matching visual cues were significantly larger when these cues matched the target colour (C+S-) than when they matched the target size (C-S+), what suggests that colour singletons might attract attention more readily than singletons defined on different visual dimensions (see also Found & Müller, 1996; Gramann, Toellner, Krummenacher, Eimer, & Müller, 2007, for evidence for special attentional processing of colour targets in visual search). Additionally, colour is known to be processed differently than other visual dimensions, i.e., the feature contrast for this dimension is computed by separate neural populations that code input from separable populations of colour analysers (for details on colour processing on the neuronal level, see Wolfe, Chun, & Friedman-Hill, 1995). It is therefore possible that the ability of partially matching visual cues to capture attention in a task-set contingent fashion in audiovisual search contexts is even more strongly reduced when these cues are defined in a dimension that is less intrinsically salient than colour.

The aim of Experiment 11 was to investigate whether attentional capture by visual singleton cues is more easily attenuated in a bimodal task context in which the task-relevant visual dimension is size. Procedures were identical to Experiment 8, except that colour singletons were now replaced by size singletons. In the unimodal Size task, participants had to discriminate the orientation of small singleton bars among medium-size distractors, and ignore search arrays with large singleton bars. The two audiovisual Size-Sound and Size-Pitch tasks were identical to

the Colour-Sound and Colour-Pitch tasks of Experiment 8, except that small bars and large bars now replaced target-colour and nontarget-colour bars as V+ and V- stimuli, respectively. In all three tasks, search arrays were preceded by spatially uninformative target-matching (small) size singleton cues. The search for audiovisually defined targets should be more effectively guided by integrated bimodal templates when these targets are defined on a dimension less salient than colour. Therefore, attentional capture by target-matching size singleton cues when both audiovisual tasks are compared to the unimodal visual Size task should now be reliably attenuated when measured by behavioural and ERP markers.

Method

Subjects

Sixteen participants took part in Experiment 11. Three were excluded due to excessive eye movements, and one due to inability to discriminate between visual targets and nontargets. The twelve remaining participants (mean age 27.9 years, age range 22–42 years, 6 females) were all right-handed and had normal or corrected vision. All gave informed consent to participate.

Stimuli, apparatus, and procedure

Experimental setup and procedures were identical to Experiment 8, except that size now replaced colour as the visual feature dimension. Cue arrays contained one smaller set of dots (0.11° x 0.11°) among five larger sets (0.17° x 0.17°). Search arrays always contained one size singleton bar (small: 0.7° x 0.17°; large: 1.9° x 0.57°) among five medium-size bars (1.1° x 0.3°). All visual stimuli were grey (CIE x/y coordinates: .308/.345; luminance: 11 cd/m^2). For all participants, small bars were designated as visual target-defining stimuli (V+) and large bars as visual nontargets (V-). The structure and trial probabilities for each of these three tasks were identical to the Colour, Colour-Sound, and Colour-Pitch tasks of Experiment 8. As size instead of colour was now used as the visual target-defining dimensions, the three tasks performed by the participants were now termed Size task (unimodal), Size-Sound task, and Size-Pitch task.

EEG recording and data analysis

These were identical to Experiment 8, except that a different time windows and onset criterion values were used for the N2pc analyses. Because the N2pc components in response to small size singleton cues were considerably smaller and emerged later than the N2pc components triggered by target-colour singleton cues in Experiments 8 and 9, N2pc mean amplitudes were now measured during the 200–310 ms interval after cue onset, and an absolute amplitude criterion of -0.4 µV was used for the jackknife-based analyses of N2pc onset latencies.

Results

Behavioural performance

Exclusion of trials with anticipatory and very slow responses led to a loss of 4% of all data. Figure 5.8 (line graphs) shows RTs for correct responses and error rates for targets at cued and uncued locations, shown separately for the three search tasks. All three tasks were performed with similar speed, $F(2,22) = 1.49$, $p = .25$. As in Experiment 8, a main effect of spatial cueing, $F(1,11) = 29.95$, $p < .001$, $\eta_p^2 = .73$, was accompanied by a two-way interaction between spatial cueing and task, $F(2,22) = 5.89$, $p < .01$, $\eta_p^2 = .35$, suggesting that spatial cueing effects differed across the visual and audiovisual search tasks. Spatial cueing effects of 38 ms found in the Size task were reduced to 22 ms and 28 ms in the Size-Sound and Size-Pitch tasks, respectively. These cueing effects were significant in all three tasks (with the smallest $F(1,11) = 15$, $p < .01$, $\eta_p^2 = .58$, in the Size-Pitch task). Planned comparisons revealed that the RTs cueing effect in the Size task was reliably larger when compared to the Size-Sound task, $F(1,11) = 10.13$, $p < .01$, $\eta_p^2 = .48$, as well as to the Size-Pitch task, $F(1,11) = 4.68$, $p < .05$, $\eta_p^2 = .3$.

As visible in Figure 5.8 (bar graphs), response errors were more frequent to targets at uncued locations relative to cued targets (9.6% vs. 5.4%; $F(1,11) = 11.81$, $p < .01$, $\eta_p^2 = .52$). Error rates were not modulated by search task and there was no task x spatial cueing interaction, both F's < 1. Participants missed 6% of all targets on Go trials. False Alarms occurred on 2% of all Nogo trials, and were not modulated by task or spatial cueing, $F(2,22) = 1.27$, $p = .3$, and $F < 1$, respectively. There was a task x spatial cueing interaction for False Alarms, $F(2,22) = 3.89$, $p < .05$, $\eta_p^2 = .26$. Pair-wise comparisons revealed that this effect was driven by a strong tendency for a difference between cued and uncued trials in the Size-Pitch task (2.2% vs. 4.1%, $F(1,11) = 4.45$, $p = .059$, $\eta_p^2 = .26$), but not in the other two tasks (smaller $p = .22$). In both audiovisual tasks,

False Alarms were more frequent on trials where target-size (small) bars appeared without concurrent tone in the Size-Sound task (V+ trials), or were accompanied by a nontarget-pitch tone in the Size-Pitch task (V+A- trials), than on trials with nontarget-size (large) bars (2.5% vs. 1.8%, and 7.7% vs. 1.8%, respectively).

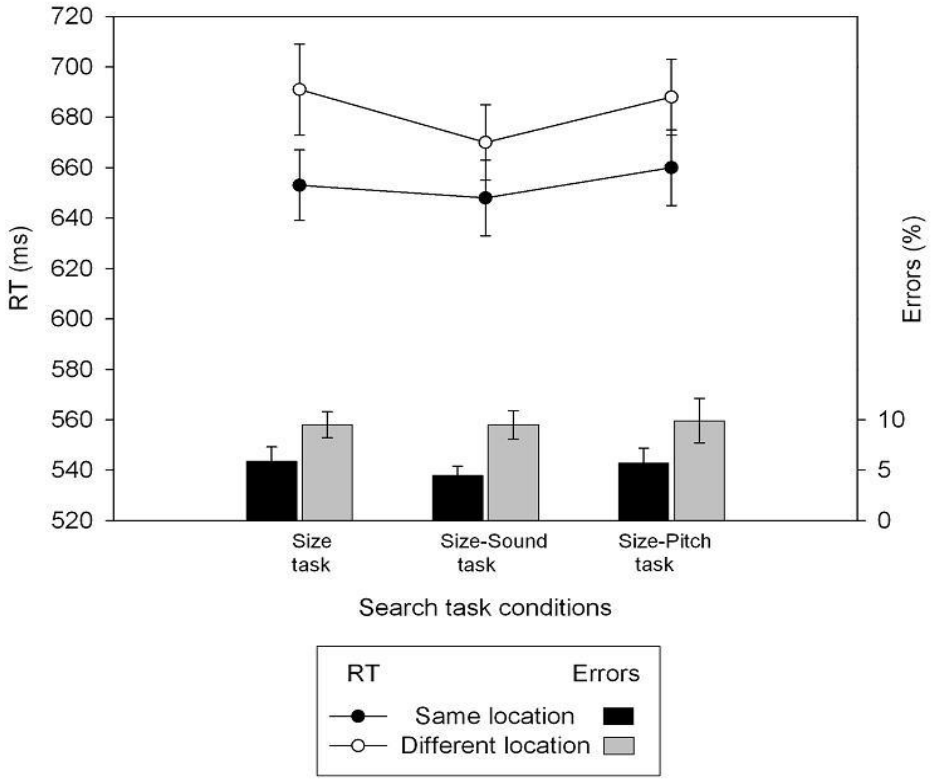

Figure 5.8. *Mean RTs (line graphs) and error rates (bar graphs) observed in Experiment 11 in response to targets at cued and uncued locations, shown separately for the Size task, the Size-Sound and the Size-Pitch task. Error bars represent standard error of the mean.*

N2pc results

Figure 5.9 (top panels) shows ERPs triggered in response to cue arrays at PO7/8 contralateral and ipsilateral to the side of the size singleton cue, separately for the three search tasks. As can also be seen in the difference waveforms obtained by subtracting the ipsilateral from contralateral ERPs (Figure 5.9, bottom panel), N2pc components were triggered by size singleton cues in all three tasks. Similar to Experiment 8, N2pc amplitudes and onset latencies differed between the unimodal visual and the two audiovisual search tasks and this was confirmed by statistical analyses.

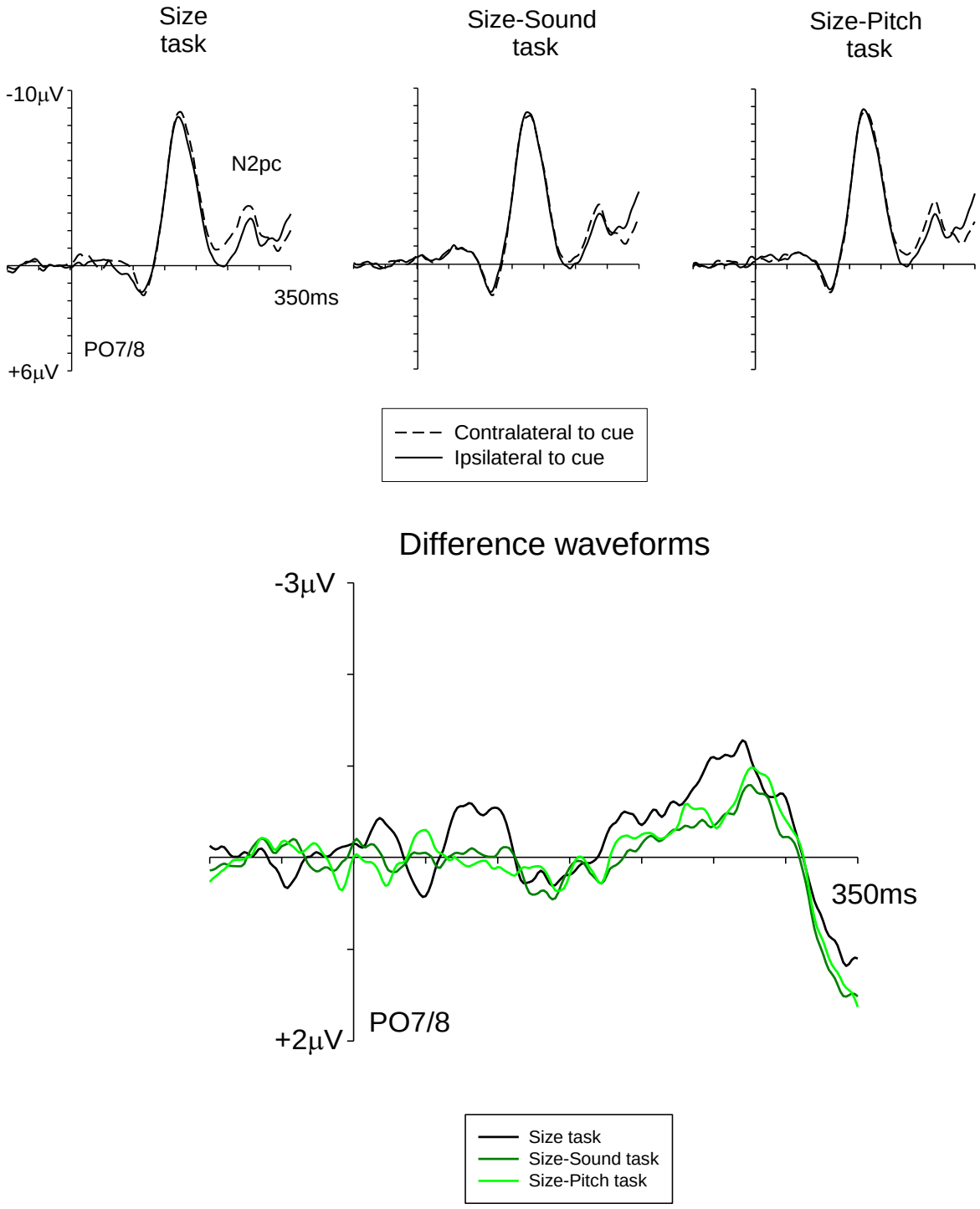

Figure 5.9. *Top panel: Grand-average ERPs measured in Experiment 11 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target- size singleton cue, separately for the Size task, the Size-Sound task, and the Size-Pitch task. Bottom panel: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the three search tasks.*

The ANOVA conducted on N2pc mean amplitudes revealed a main effect of contralaterality, $F(1,11) = 15.03$, $p < .01$, $\eta_p^2 = .58$, thus suggesting reliable N2pc components in response to singleton size cues in all three search tasks. This was confirmed by planned comparisons, with the smallest $F(1,11) = 3.41$, $p < .05$, $\eta_p^2 = .24$, found for the cue-elicited N2pc in the Size-Sound search task. Most importantly, there was an interaction between contralaterality and task, $F(2,22) = 7.4$, $p < .01$, $\eta_p^2 = .4$, which demonstrated that N2pc amplitudes varied across tasks. In a marked contrast to Experiment 8, planned contrasts revealed that the cue-elicited N2pc component in the Size task was reliably larger when compared both to the Size-Sound task, $F(1,11) = 6.28$, $p < .05$, $\eta_p^2 = .36$, as well as to the Size-Pitch task, $F(1,11) = 10.05$, $p < .01$, $\eta_p^2 = .48$. Furthermore, N2pc onset latency in the Size task (203 ms) was significantly earlier when compared both to the Size-Sound task (255 ms; $t_c(11) = 2.64$, $p < .05$) and to the Size-Pitch task (226 ms; $t_c(11) = 4.32$, $p < .01$).

Discussion

The aim of Experiment 11 was to investigate whether defining the audiovisual targets in the visual dimension of size will produce more robust effects of bimodal search templates on attentional object selection, with reductions of task-set contingent capture observed irrespective of whether the partially-matching cues are perceptually similar to nontargets or not. The observed results demonstrated that in the contexts, where size, rather than colour, was the task-relevant visual dimension, the ability of visual singleton cues to attract attention is indeed reliably reduced in both bimodal search tasks. Similarly to target-colour singleton cues in Experiments 8, target-matching small singleton cues elicited significantly smaller RT spatial cueing effects in the two audiovisual tasks than in the unimodal visual Size task. However, in contrast to Experiment 8, the N2pc component to size singleton cues was now attenuated and delayed in both audiovisual tasks relative to the unimodal task. If the salience of the visual dimension did not play a role in the guidance of search for audiovisually defined targets by integrated bimodal target templates, a pattern of behavioural and electrophysiological attentional capture effects very similar to Experiment 8 should have been observed in the present experiment.

The role of the visual dimension defining the audiovisual target in the control of selection by integrated bimodal object templates can also be visible in a few notable differences in the capture effects obtained for size-defined singletons in Experiment 11, as compared to colour singletons in Experiments 8 and 9. First, in Experiment 8, behavioural spatial cueing effects were completely eliminated in the Colour-Sound task. In Experiment 11, they were significantly

reduced in the Size-Sound task as compared to the unimodal Size task, but remained reliably present. The fact that search for size-defined targets was more difficult than search for colour-defined targets in Experiments 8 and 9 (as reflected by longer RTs and higher error rates in Experiment 11) may have been responsible for this difference. Reduced forward masking by size singleton cues could also have contributed to the residual spatial cueing effects in the Size-Sound task: Discriminating the orientation of small target singleton bars is likely to have been easier when they were presented at a cued location (i.e., a location previously occupied by the smallest element of the cue array) relative to uncued locations (i.e., locations previously occupied by a larger cue array element; see also Kiss & Eimer, 2011). What is important is that both task difficulty and forward masking were factors that remained constant across all three tasks in Experiment 11. Thus, the reduction of spatial cueing effects for partly-matching cues in the audiovisual as compared to unimodal task contexts can still be attributed to the influence of bimodal attentional templates.

The second difference concerns the magnitude of cue-elicited N2pc amplitudes in Experiment 11 relative to Experiment 8 and 9. The N2pc triggered by size singleton cues in Experiment 11 was much smaller than the N2pc amplitudes elicited in response to colour singleton cues in the previous experiments. This pattern of results is in line with previous ERP studies of attentional capture by visual feature singletons, which separately found larger N2pc components for colour singletons as compared to singletons that were defined in another dimension such as size (Kiss & Eimer, 2011) or shape (Seiss, Kiss, & Eimer, 2009). It is possible that this difference in N2pc amplitudes between colour singletons and other-dimension feature singletons reflects the stronger bottom-up salience of feature contrasts in the colour domain. Notably, despite the N2pc amplitude differences between colour and size singleton cues found across Experiments 8, 9 and 10, the amplitude reductions and onset latency delays observed in audiovisual as compared to unimodal visual task contexts were very similar for both types of cues.

The important novel insight provided by Experiment 11 into the control of task-set contingent attentional capture by integrated object templates in search for multi-feature bimodal targets is that the visual dimension on which audiovisual targets are defined plays an important role in this control. The ability of target-matching distractors to initially attract and subsequently hold attention in their location, as indexed by the N2pc component and behavioural spatial cueing effects, respectively, is more effectively reduced during audiovisual search that takes place in contexts where the targets are defined on a visual dimension that is intrinsically less salient than colour. Compared to colour-pitch conjunctions, in audiovisual size-pitch conjunction search tasks

the ability of partly-matching distractors to capture attention is reduced irrespective of whether they are perceptually similar to nontargets or not. Lastly, the current findings demonstrate that guidance of attentional selection by integrated bimodal templates generalises across target-defining visual feature dimensions, and is clearly not just characteristic of task sets that involve colour.

Discussion of Chapter 5

The ability to create representations of task-relevant information determines effective behaviour and cognitive functioning in real-life environments (Desimone & Duncan, 1995; Duncan et al., 1997). Existing studies of visual attention or multisensory processing (Iordanescu et al., 2008, 2010; Kiss et al., 2013) have left unanswered an important question concerning whether spatial selection can be guided in a flexible task-dependent fashion towards bimodal compared to unimodal objects when the former match both features of a target that is defined across modalities (e.g., a colour-pitch conjunction). The aim of Chapter 5 was to assess whether search for audiovisual target objects is guided by integrated bimodal object templates or independent within-modality representations of task-relevant information. For this purpose, RTs spatial cueing effects and N2pc components triggered by unimodal target-matching visual cues were compared across tasks where targets were defined by a visual feature alone or a combination of visual and auditory features. Across Experiments 8 through 11, behavioural and electrophysiological capture effects elicited by unimodal target-similar cues were typically strongly attenuated, and sometimes even completely eliminated, in audiovisual versus visual search contexts. These results converged to provide the first evidence for the importance of integrated bimodal object templates in the top-down control of search for bimodal targets.

If attentional selection during search for multi-feature multimodal targets was controlled by fully independent within-modality target templates, with attentional visual and auditory target feature representations operating in a strictly independent fashion, attentional capture effects triggered by task-set matching visual feature singleton cues should have been similar across visual and audiovisual search contexts. However, what needs to be noted is that if search for audiovisual targets was guided by templates where visual and auditory target features are *fully* integrated into a single object representation, one would expect attentional capture by task-set matching visual cues to be absent in the audiovisual tasks, as attention would only be allocated to fully template-matching objects. In fact, although behavioural and electrophysiological markers of attentional capture by visual cues were shown to be reliably attenuated during audiovisual

search in three out of four of the reported experiments, these capture effects were clearly not completely eliminated. On the one hand, these findings are inconsistent with attentional guidance by fully integrated bimodal objects, and suggest instead that attentional control of audiovisual search retains some modality-specific aspects. On the other hand, the fact that spatial cueing effects and cue-elicited N2pc components were consistently reduced in audiovisual as compared to unimodal visual task contexts demonstrates that bimodal attentional templates play an important role in the guidance of search for audiovisual targets.

Care is warranted when concluding that the current findings suggest that guidance of attentional selection by integrated objects representations is less effective in case of audiovisual compared to purely visual templates. In the only directly comparable study of Kiss et al. (2013; Experiment 1), partly-matching visual cues triggered no reliable RT spatial cueing effects, indicative of the inability of such cues to capture attention during colour-size search. In the audiovisual search contexts employed in the experiments reported in the present chapter, a complete absence of reliable behavioural spatial cueing effects was observed only in one case (Colour-Sound task in Experiment 8). This pattern of results does not necessarily indicate a weaker control of task-set contingent capture by audiovisual as compared to visual templates. Instead, these findings point towards a potentially crucial role of perceptual similarity between the target-matching cues and stimuli explicitly defined as nontargets in a given task context. Nonrealiable behavioural spatial cueing effects were found in both the Colour-Sound task in Experiment 8 and in the study of Kiss et al. (2013). What is important, in both of these task contexts cues were perceptually similar to nontargets in the search array. In circumstances in which such similarity was absent, reliable reductions of behavioural and ERP markers of attentional capture were found in contexts where the target-defining pitch was strongly indicative of target presence (Experiment 9) or where bimodal targets were defined on visual dimensions less intrinsically salient than colour (Experiment 11). In neither of these situations, however, capture was completely eliminated. This indicates that disengagement from the location of the target-similar distractor, i.e., the selection stage of task-set contingent capture that might be predominantly controlled by integrated object representations (Kiss et al., 2013), is strongly modulated by whether such cue stimuli possess attributes that are shared by stimuli explicitly defined as nontargets. In circumstances in which the irrelevant cues do not match these nontarget stimuli, the ability of audiovisual task sets to initiate disengagement from these cues is weaker, although not completely eliminated. More research is required to determine whether similar conclusions hold for unimodal visual search contexts.

Research into how attentional object selection is controlled during search for targets defined by (cross-modal and within-modal) feature-conjunction targets might demonstrate a somewhat different hierarchy between bottom-up and top-down mechanisms than one found for single-feature search (cf., Eimer et al., 2009; Folk et al., 1992; Lamy et al., 2003, 2004). The findings reported in Chapter 5 indicate that bottom-up salience may play an important role in modulating the ability of target-matching irrelevant objects to capture attention in search for multi-feature targets. Namely, a complete elimination of spatial cueing effects in audiovisual contexts was shown only for feature singleton cues (Colour-Sound task in Experiment 8; see also Kiss et al., 2013), but not for heterogeneous cues (Colour-Sound task in Experiment 10). However, more work is required to assess the importance of bottom-up factors in the control of contingent capture during search for multi-feature target objects. It will be important for future work to establish whether there are other important factors, such as the dissimilarity between the cues and nontarget search array stimuli, that may have prevented the visual cues in Experiment 10 from being subject to active top-down inhibition. Additionally, in systematic research into the role of differences in the intrinsic salience between various visual dimensions in the top-down guidance of attentional object selection by integrated object templates, it might be necessary to control for the level of general task difficulty that might be characteristic of search for feature-conjunction targets defined in specific visual dimensions. In spite of this, the current findings on top-down control of selection by feature-conjunction templates have the potential to enrich our current knowledge in respect to the relative importance of salience- and goal-based mechanisms in the control of attentional objects selection in real-life environments.

On a related note, it remains to be established where multimodal object templates are created and maintained in the brain. Such attentional templates may be represented in multimodal brain regions, such as FC, STS, or LIP, with the specific region likely dependent on familiarity and semantic congruence between features from different modalities (Hein et al., 2007; Naumer et al., 2009; but see also Taylor et al., 2006). What is worth noting, bimodal templates were shown to control neural processing in extrastriate cortical areas that are regarded traditionally as modality-specific areas. The N2pc component is as a modality-specific visual component which originates primarily, but not entirely, in the extrastriate ventral visual cortex (Hopf et al., 2000), and is elicited by task-relevant visual objects. The findings from experiments reported in Chapter 5 demonstrated that the N2pc component in response to task-set matching visual singleton cues can be attenuated and delayed during audiovisual versus visual search, thus providing the first evidence that bimodal templates, i.e., integration of representations of target-defining features

from different modalities, can modulate spatially selective processing in modality-specific visual areas.

How could this top-down control over attentional selection by bimodal attentional templates be implemented? In the Guided Search model of visual object selection (Wolfe 1994, 2007), the input from separate feature channels to the central salience map is weighted according to the task relevance of specific features. As target features that are relevant at a given point in time are weighted high, they create a strong spatial bias in the activity profile on the salience map, which results in preferential attentional selection of target objects. It is possible that during search for audiovisual targets, target-defining visual attributes also receive a positive weighting, but that these top-down weights and the resulting spatial bias in favour of task-set matching visual features are reduced when targets are defined across sensory modalities, when compared to unimodal single-feature targets. To provide an example, during search for red singleton bars that are accompanied by high-pitch tones, feature channels coding red objects will be less strongly weighted (and thus have smaller impact on the activity profile of the salience map) relative to unimodal search for red bars. Such a mechanism would explain why capture by colour (and size) singletons was reduced across all the experiments reported in the Chapter 5. In contexts where unimodal feature singletons serve as to-be-ignored nontargets (i.e., Colour-Sound task in Experiment 8), any top-down biases in favour of such target-matching feature singletons might be further reduced.

Additionally, in the experiments reported in this chapter, but not in the study of Kiss et al. (2013), reliable N2pc amplitude reductions in audiovisual task contexts were always accompanied by N2pc onset delays. The mechanisms controlling task-set contingent capture during search for feature-conjunction targets have just begun to be investigated and it is therefore unclear what is responsible for this N2pc latency shift. While these onset delays may reflect an effect of top-down control of contingent capture by integrated object templates that is specific only to audiovisual search contexts, more work is required to establish whether this effect may not instead be driven by specific features of the task design.

Overall, the findings reported in Chapter 5 provide novel evidence that the ability to effectively control attentional capture by objects partly matching multi-feature targets generalises across within- and multi-modal search contexts. Reduced behavioural and electrophysiological capture effects in audiovisual versus visual search demonstrate that flexibility of top-down mechanisms to control search for objects that are defined by an arbitrary conjunction of features is not restricted to within-modal visual contexts, in spite of the dominance of the visual representations in the object identification and control of spatial behaviour in the external

environment (Goodale & Milner, 1992; Mishkin & Ungerleider, 1982). Thus, the current findings provide another line of evidence for the tenets of the biased competition model (Desimone & Duncan, 1995): The flexibility of top-down biasing mechanisms in creating representations of task-relevant information with the purpose to bias neural processing and control of behaviour (Duncan et al., 1997; Duncan, 2010) extends to multimodal target objects, which are typical for real-life environments. Further research is required to establish whether the retention of the ability of target-similar cues to capture attention shown in the experiments reported in Chapter 5 reflects a more general characteristic of top-down control by object templates. The present results provide an important extension also to another major theory of visual attention, i.e., the guided search model (Wolfe, 1994, 2007): The activity on the central salience map, originally assumed to be involved purely in allocation of attention to objects in *visual* space, is controlled by bimodal search templates. As an electrophysiological marker of the current activity profile on this map, the N2pc component proved to be useful in providing intriguing novel insights into the relative hierarchy of top-down and bottom-up factors affecting contingent attentional capture during multi-feature object search.

Lastly, an important implication of the present findings is that visual selection can be biased towards visual objects accompanied by stimuli in other modalities on the basis of mechanisms other than just their heightened bottom-up salience (Matusz & Eimer, 2011), i.e., their increased task-relevance. Research into the interplay between selective attention and multisensory integration has focused almost exclusively on the salience-based mechanism (van der Burg et al, 2008a, 2008b, 2011, 2012; Vroomen & de Gelder, 2000), and has rarely addressed the role of cross-modal integrative processes as a source of a top-down bias in attentional selection. This gap is particularly notable if one considers the substantial interest that the feature-based top-down control of visual attention selection has received in the past two decades (Eimer et al., 2008, 2009, 2010; Eimer et al., 2011; Folk et al., 1992, 1998; Lamy et al., 2003, 2004; Theeuwes, 1991, 1994, 2010; Yantis & Egeth, 1999). The two studies existing in the cross-modal literature have shown that visual search for naturalistic audiovisually defined objects can be facilitated by presence of a semantically congruent (and thus template matching), feature in another modality (Iordanescu et al., 2008, 2010). However, in contrast to the experiments described in Chapter 5, top-down task set was not well controlled in these studies. Therefore, they cannot be treated as strong evidence for the role of multisensory integration as a source of top-down bias in attentional selection via integrated object templates.

Overall, the results from the experiments reported in Chapter 5 provide novel evidence that search for audiovisual target objects is not exclusively controlled by independently operating

modality-specific representations of target-defining features. They demonstrated also that early stages of attentional selectivity in extrastriate visual cortex can be already modulated by bimodal attentional templates. Further work is required to provide a better understanding of how control of spatial selection by integrated object templates differs between cross-modal and within-modal search contexts.

Chapter 6. Conclusions

The theoretical and methodological advancements made in the last two decades have paved the way for questions concerning the mechanisms underlying different types of interactions between selective spatial attention and multisensory integration (see Koelewijn et al., 2010; Talsma et al., 2010 for reviews). In the present thesis, two types of interaction between multisensory integration and spatial attention that can occur in multi-stimulus contexts have been discussed. The first interaction pertains to multisensory enhancement of attentional capture via an increase of bottom-up salience of visual objects paired with non-visual signals (Matusz & Eimer, 2011; Olivers & van der Burg, 2008). The second interaction concerns the presence of a top-down bias in spatial selective attention towards bimodal compared to unimodal objects in contexts where bimodal objects match both features of an audiovisually defined target (cf., Iordanesco et al., 2010).

As highlighted in Chapter 1, the mechanisms that support the first type of interaction have received some interest in the last decade, and the research presented in Chapters 2 to 4 has substantially contributed to their understanding. A now-classical study in the area of visual attention has demonstrated that salient objects fail to capture attention if they do not share task-relevant features (Folk et al., 1992). Since then, converging evidence has been provided for the contingence of involuntary shifts of attention on goal-based mechanisms, where irrelevant stimuli are selected in contexts in which they match the features of the current target and not when they are merely distinctive from their visual background but do not share any of the task-relevant features (Eimer et al., 2008, 2010; Folk et al., 1992, 1998; Hickey et al., 2008; Lamy et al., 2003, 2004; Sawaki & Luck, 2010). Notably, research that was carried out independently has demonstrated that the bottom-up salience of visual objects might also be increased by multisensory integration. Behavioural and neural responses to visual objects were shown to be enhanced in situations where the visual objects temporally coincided with signals in another modality (e.g., Giard & Peronnet, 1997; Stein et al., 1996). These preliminary findings have motivated studies that provided evidence that visual attention can be oriented more strongly towards visual objects accompanied by irrelevant and uninformative tones because such pairings tend to be automatically integrated at sensory-perceptual levels of cortical processing into salient multi-modal objects (Olivers & van der Burg, 2008; Vroomen & de Gelder, 2000). However, it remained unclear whether audiovisual synchrony could also play an important role in multi-stimulus contexts by enhancing selection of visual objects via a bottom-up mechanism that is not contingent on top-down attentional control (cf., van der Burg et al., 2008a, 2008b, 2011). Thus,

research presented in Chapters 2 to 4 of this thesis focused on two major questions: Does audiovisual salience reliably enhance the ability of visual objects to capture attention in all contexts in which multiple simultaneous objects compete for selection? And, critically, can visual objects, whose bottom-up salience was increased by audiovisual synchrony, be selected through a mechanism that operates independently of top-down, goal-based mechanisms? Section 6.1 summarises the findings and conclusions with respect to the role of audiovisual salience in biasing visual object selection.

The mechanisms that underlie the second type of interaction have received virtually no attention in the past, and the present thesis has made a critical contribution to their understanding. Very little is known about the mechanisms of top-down attentional control that support search for targets defined across modalities (cf., Eimer et al., 2002, for findings on how spatial attention is controlled across modalities). It has been previously demonstrated that search for objects is facilitated in cases where objects contain both visual and auditory, rather than just the visual target-defining feature (Iordanesco et al., 2008, 2010). However, these findings can be accommodated by guidance of attention by separate features (Treisman & Gelade, 1980; Wolfe, 2007) as well as by integrated object templates (Duncan et al., 1997), and, thus, do not provide direct insights into the attentional control mechanism underlying the top-down bias towards bimodal objects. Research in visual attention has provided mounting evidence for the flexibility of top-down control mechanisms in creating representations of task-relevant information (Corbetta & Shulman, 2002; Desimone & Duncan, 1995). In line with this, recent findings (Kiss et al., 2013) have demonstrated that search for multi-feature targets can be controlled by integrated object templates. However, these findings cannot rule out the possibility that such form of attentional guidance is limited to within-modal visual targets: The dominating role of vision in localisation and object identification (Welch & Warren, 1980) suggests that selection might be controlled by modality-specific representations of task-relevant features. The need to better understand how attentional selection is controlled during search for targets that are defined across modalities has motivated the research reported in Chapter 5 of this thesis. Two major questions were addressed: Can search for targets defined by conjunctions of visual and auditory features be controlled by fully integrated bimodal object representations? Are there factors that facilitate guidance of attention based on such integrated audiovisual templates? The contribution of the findings reported in Chapter 5 to the research on the mechanisms of attentional control during search for objects defined across modalities is summarised in Section 6.2.

6.1. Mechanisms underlying salience-based biases in visual selection towards synchronous audiovisual stimuli

The investigations into the role of audiovisual salience in the selection of visual objects in space have yielded two important sets of findings. On the one hand, new direct evidence has been provided for audiovisual salience as a source of bias in visual selection that operates independently of top-down task set (6.1.1). On the other hand, a novel insight was provided into the importance of within-modal salience for the ability of irrelevant audiovisual synchronous objects to trigger a stronger bias in visual selection relative to unimodal visual objects (6.1.2).

6.1.1. Audiovisual synchrony as a mechanism of bottom-up bias in visual object selection

In Experiments 1 to 6, the spatial cueing paradigm (Folk et al., 1992) was adapted for a multisensory context, and behavioural and ERP responses were measured to investigate whether audiovisual synchrony can enhance the ability of task-irrelevant colour cues to capture attention by increasing their bottom-up salience. On the one hand, research on visual attention (Bacon & Egeth, 1994; Eimer et al., 2009, 2010; Folk et al., 1992, 1998; Lamy et al., 2003, 2004) has suggested that visual distractors capture attention, as measured by behavioural cueing effects and the N2pc component, only in contexts in which they match features of the current target. On the other hand, converging neurophysiological and behavioural evidence for automatic integration of temporally coincident signals from different modalities into salient multimodal objects at low levels of the cortical hierarchy (see Cappe et al., 2009, Driver & Noesselt, 2008; Koelewijn et al., 2010) predicts that the ability of visual objects to capture visuo-spatial attention should be enhanced in contexts where these visual objects are accompanied by non-informative tones, and this enhancement should be independent of their task-relevance.

In Experiment 1, search arrays contained a colour singleton bar presented among five grey distractor bars, and were preceded by colour-change singleton cues. Critically, the spatially non-predictive colour changes were accompanied on 50% of all trials by spatially diffuse task-irrelevant tones. Target bars could have one of two possible colours that were randomly intermixed within each block. The colour-change cue could match one of these colours or have a third, nontarget colour. As selection in this experiment was controlled by local feature contrasts (‘singleton-detection mode’; Bacon & Egeth, 1994; see also Eimer & Kiss, 2010), larger RT spatial cueing effects that were triggered by colour cues on trials on which the cues were paired

with tones are in line with multisensory integration enhancing the ability of visual objects to capture attention by increasing their bottom-up salience. These enhancements could not be explained by tone-induced alertness: In a task context designed to maximise tone-induced alerting effects on performance, i.e., when tones were presented concurrently with the onset of the base array preceding the cue array, no enlargement of cueing effects was observed as a function of tone presence (Experiment 2). However, because participants adopted a singleton-detection mode in Experiment 1, the question whether these observed enhancements are contingent on top-down task sets could not be addressed. Hence, to provide a more direct test of the bottom-up nature of the mechanism by which audiovisual synchrony creates a bias in visual object selection, target bars of one predefined colour were presented among differently coloured distractor bars in Experiment 5, thus forcing participants to adopt a feature-specific task set. Importantly, as the colour-change cues could match the target colour or share a nontarget colour, it was now possible to assess whether audiovisually induced enhancements of capture effects were contingent on top-down attentional control settings. In line with the assumption that participants would have now adopted a colour-specific feature search mode (Bacon & Egeth, 1994), only target-colour cues, but not nontarget-colour cues, elicited reliable spatial cueing effects. Critically, the enlargements of capture effects by colour-change cues on trials where the cues were paired with tones were found irrespective of whether cues shared the target colour or not, thus providing the first direct evidence for the bottom-up nature of the mechanisms by which multisensory integration creates a bias in visual object selection.

When ERP responses to colour cues on tone-present and tone-absent trials were compared (Experiment 6), enhanced N2pc amplitudes were found, demonstrating that at the neural level, audiovisual and visual salience enhance visual selection bias via a similar mechanism (cf., Eimer et al., 2009). However, N2pc enhancements were observed only for singleton, but not heterogeneous, colour cues (see Section 6.1.2 for more details). Multisensory modulations of visual object selection, as indexed by the N2pc component, were subsequently replicated in a task context in which the search target was defined as a conjunction of visual and auditory features (Experiment 7). Tones were now presented together with colour-defined bars in the search arrays and could match the target or nontarget pitch. N2pc enhancements of similar size were found to target-colour and nontarget-colour bars when these were accompanied by task-set irrelevant tones, what provided strong evidence that audiovisual synchrony enhances selection of objects in space independently of top-down attentional control settings.

6.1.2. The role of within-modal salience in the audiovisual enhancements of bottom-up selection bias in vision

Further evidence for the bottom-up nature of the mechanisms by which audiovisual salience enhances attentional capture in vision was provided by the experiments reported in Chapters 2 to 4 that investigated the factors that modulate these enhancements. The relative salience of integrated audiovisual cues was found to be important for the ability of bimodal cues to attract involuntary attention more strongly than purely visual cues. When paired with lower-intensity tones, only colour-change cues presented against a homogeneous, but not heterogeneous, background elicited larger behavioural spatial cueing effects on tone-present compared to tone-absent trials (Experiment 1 vs. Experiment 4). Critically, reliably enhanced capture effects as a function of audiovisual synchrony were also observed in response to lower-salience heterogeneous colour cues, but only when higher-intensity sounds were used (Experiment 3 vs. Experiment 5). This contrasts with the results demonstrating that in contexts where heterogeneous colour cues were paired with lower-intensity tones, tone presence did not enhance spatial cueing effects, irrespective of whether the participants adopted a high-selectivity feature-search mode or a low-selectivity singleton-detection mode (Experiment 3 vs. Experiment 4; Bacon & Egeth, 1994). The importance of within-modal salience for the presence of an audiovisually induced bottom-up selection bias in vision was further substantiated by electrophysiological results: In a task context where heterogeneous and singleton colour failed to trigger enlarged behavioural capture effects on tone-present versus tone-absent trials (Experiment 6), audiovisual singleton colour cues still elicited reliably enhanced N2pc amplitudes.

6.2. Mechanisms underlying the top-down control of spatial selection by integrated audiovisual object templates

The experiments that are reported in Chapter 5 were aimed at investigating whether the preferential selection of bimodal objects during search for objects defined across modalities is mediated by integrated object templates. The findings from these experiments will be discussed in two separate sections. First, novel evidence was provided in support of the idea that in contexts in which search targets are defined as conjunctions of features from different modalities, attention can be guided by integrated object templates (6.2.1). Second, the factors that play an important role in attentional guidance during search for bimodally defined targets were revealed (6.2.2).

6.2.1. Top-down control of spatial selection by integrated audiovisual object templates

To investigate the mechanisms underlying the top-down control of search for targets that are defined as conjunctions of features from different modalities, the experiments reported in Chapter 5 employed another variant of the Folk et al.'s (1992) cueing paradigm: Attentional capture triggered by unimodal target-matching visual cues, as indexed by behavioural and ERP measures, was compared across search tasks where targets were defined by a visual feature alone or by visual and auditory features simultaneously. In the visual task, participants had to respond to bars defined by one pre-specified feature (e.g., red bars), and ignore bars with a nontarget feature (e.g., blue bars). In the audiovisual tasks, target trials were defined by a combination of visual and auditory features (e.g., red bars accompanied by high-pitch tones), and nontarget trials were defined by presence of nontarget features appearing in one or both modalities. Across Experiments 8 to 11, RT spatial cueing effects triggered by unimodal target-matching cues were reduced during search for audiovisual as compared to unimodal visual targets. In all experiments in which the task-set matching visual cues were singletons (all except for Experiment 10), the cue-elicited N2pc component was attenuated and delayed in onset during audiovisual search. This converging pattern of behavioural and electrophysiological reductions of capture effects is inconsistent with the account proposing that search for targets defined as conjunctions of features from different modalities is guided by separate modality-specific representations of task-relevant information. Instead, the current findings provide support for the idea that search for multimodal objects can be guided by integrated object templates. Integrated attentional templates will reduce or eliminate the ability of distractors that match only one of the target-defining features of a multi-modal target to capture attention. Importantly, these reductions were observed irrespective of whether targets were defined in the dimension of colour (Experiments 8 and 9) or size (Experiment 11), thus highlighting the generalisability of this top-down control mechanism.

6.2.2. Top-down and bottom-up factors modulating control of spatial selection in audiovisual search contexts

To provide an insight into the mechanisms supporting the guidance of spatial selection during audiovisual search by integrated object representations versus separate modality-specific representations of task-relevant information, the role of both top-down and bottom-up factors was explored. In the experiments reported in Chapter 5 (except for Experiment 9), two bimodal search

tasks were employed, originally designed to assess how demands associated with processing of target-pitch tone (detection versus discrimination) affected guidance by audiovisual templates: In one of the audiovisual tasks, target-matching visual bars in the search array were always presented alone. In the other audiovisual task, the same target-matching bars were accompanied by nontarget-pitch tones. Participants maintained the same audiovisual task set across the two tasks in all three experiments, as indicated by similar RTs and error rates. In a task context involving search for colour-pitch targets (Experiment 8), reliable reductions of behavioural and ERP capture effects were found only in the bimodal task, where target-colour bars were presented without tones on nontarget trials (Colour-Sound task), that is, where the perceptual features of these nontarget arrays were similar to those of the cue arrays. This suggested that top-down inhibitory processes, associated with perceptual similarity between cue and nontarget arrays, might play an important role in supporting top-down control over target-similar distractors during search for multi-modal targets (cf., Kiss et al., 2013). Importantly, attentional capture by distractors matching only one of the target-defining features can be reduced during audiovisual search even in tasks where these distractors are not similar to stimuli explicitly defined as nontargets: Behavioural and ERP capture effects triggered by target-colour cue were reliably attenuated in a context, where the target-defining pitch was strongly indicative of the presence of an audiovisual target (Experiment 9). These results also indicated that weights assigned to inputs from the visual channels can be flexibly adjusted to control search in line with specific task demands even in audiovisual search task contexts (cf., Wolfe, 2007). The current findings also highlighted the important role of independently operating input channels during audiovisual search: Unimodal target-colour cues presented against a multi-coloured background showed reduced behavioural capture effects during audiovisual versus visual search, but these cues triggered similar N2pc amplitudes across these two search contexts (Experiment 10). This finding suggests that the initial stage of selection might be controlled predominantly, albeit not entirely, by input from independent feature channels (see also Kiss et al, 2013, for similar arguments). Overall, the results presented in this section provide a new perspective on the interactions between independent modality-specific and genuinely multimodal mechanisms that are involved in the control of attentional selection during search for target objects that are defined in different sensory modalities.

6.3. Future directions

A number of important unresolved issues with respect to mechanisms controlling attentional selection of objects in multisensory environments were raised by the findings reported in the present thesis. Even though audiovisual salience was shown to enhance the ability of irrelevant visual objects to capture attention independently of top-down task set, this effect might still be modulated by temporal attention. When salient irrelevant audiovisual objects are presented at known task-irrelevant points in time, they can initially receive a competitive advantage at early stages of perceptual processing, but their ability to maintain attention at their location might be reduced when compared to audiovisual objects that occur at a potentially task-relevant time. In real-life environments, potentially relevant objects are usually not only multi-modal, but also temporally unpredictable. The interaction between audiovisual enhancements of attentional capture and temporal attention will need to be investigated in future studies.

Another important question is whether top-down control based on bimodal attentional templates operates analogously to the way that unimodal visual attention is controlled in circumstances, where targets are defined by a combination of features from different visual dimensions. For example, the pattern of behavioural and electrophysiological capture effects observed in Experiment 8 during search for audiovisually defined targets was similar to the pattern found during search for visual targets defined by a combination of colour and size (Kiss et al., 2013). Does this similarity imply that the mechanisms underlying top-down control by integrated object representations are similar irrespective of whether targets are defined within or across sensory modalities? This account would be in line with the supposed supramodal nature of attentional control mechanisms that was previously studied primarily in the context of selection of locations, rather than objects, in space (cf., Eimer et al., 2002; Farah et al., 1989). On the other hand, one could assume that top-down attentional control mechanisms might be implemented more efficiently when targets are defined solely within a single sensory modality, e.g., by directly adjusting weights on a hypothetical visual salience map in a task-set dependent fashion (e.g., Wolfe, 2007) without further input from higher-order multimodal attentional control regions. The question whether there are systematic differences in the efficiency of selecting targets that are defined within and across sensory modalities awaits further clarification. In this context, N2pc latency measures might potentially be important.

6.4. Summary and implications

The research presented in this thesis was born out of the theoretical and methodological advancements that were made in the past thirty years and which have enabled questions into how mechanisms controlling spatial selection of objects that are defined across modalities differ from unimodal visual or auditory control mechanisms. The findings reported here have provided valuable insights into the mechanisms by which multisensory integration can create a bottom-up as well as a top-down bias in spatial object selection.

First, audiovisual synchrony was shown to enhance the ability to capture attention of irrelevant visual objects paired with non-visual signals that appear in multi-stimulus environments by increasing their bottom-up salience (Chapters 2 to 4). While both visual and audiovisual forms of perceptual salience modulate spatial selection by enhancing the neural responses that are triggered by objects, audiovisual salience enhances attentional selection independently of the top-down task set. Second, search for targets that are defined by arbitrary pairings of visual and auditory features is guided by integrated bimodal object templates (Chapter 5). However, to some extent, it might also be controlled by modality-specific representations of task-relevant information, as irrelevant visual objects that match current target-defining features retain some ability to capture attention.

The present thesis has contributed to the current knowledge on attentional control mechanisms by demonstrating that spatial selection can be frequently biased towards multimodal versus unimodal objects. One mechanism that underlies this bias operates in a bottom-up salience-based fashion; another is contingent on a top-down multimodal task set. More generally, the present thesis demonstrates that comprehensive models of how selective attention is controlled in naturalistic environments cannot just be based on unimodal research, but also need to take into account how objects and events are selected in real-world contexts, which are multisensory by nature.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1339–45.

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, *23*, 3–38.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*, 839–43.

Andersen, R. A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*, 1421–1428.

Andersen, R. A., Snyder, L. H., & Bradley, D. C. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, *20*, 303–330.

Arrington, C. M., Carr, T. H., Mayer, A. R., & Rao, S. M. (2000). Neural mechanisms of visual attention: object-based selection of a region in space. *Journal of Cognitive Neuroscience*, *12*(2), 106–117.

Assad, J. A., & Maunsell, J. H. R. (1995). Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature*, *373*, 518–521.

Astle, D. E., Nobre, A. C., & Scerif, G. (2010). Subliminally presented and stored objects capture spatial attention. *The Journal of Neuroscience*, *30*(10), 3567–3571.

Avillac, M., Ben Hamed, S., & Duhamel, J.-R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *The Journal of Neuroscience*, *27*(8), 1922–1932.

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 834–846.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception*, *55*(5), 485–496.

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 948–961.

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8*, 234–238.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Volume 8* (pp. 47–90). New York, NY: Academic Press Inc.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*(3), 99–102.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, *17*(3), 377–391.

Barth, D. S., Goldberg, N., Brett, B., & Di, S. (1995). The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Research*, *678*, 177–190.

Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M. J., Hénaff, M. A., Michèl, F., & Barton, J. J. (2001). Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron*, *32*(6), 985–995.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*(11), 1190–1192.

Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*(8), 1110–1116.

Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision*, *7*, 1–12. doi:10.1167/7.2.19.

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*(10), 1154–1165.

Bell, A. H., Meredith, M. A., Van Opstal, A. J., Munoz, D. P., & Andrew, H. (2005). Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *Journal of Neurophysiology*, *93*, 3659–3673.

Bernstein, I. H., Clark, M. H., & Edelstein, B. A. (1969). Effects of an auditory signal on visual reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *80*, 567–569.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion of auditory–visual spatial discordance. *Perception & Psychophysics*, *29*, 578–584.

Besle, J., Fort, A., & Giard, M.-H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing*, *5*(3), 189–192.

Bishop, S. J. (2008). Neural mechanisms underlying selective attention to threat. *Annals of the New York Academy of Sciences*, *152*, 141–152.

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). " Acoustical vision " of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*, 273–282.

Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologica*, *50*, 253–290.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*(2), 369–384.

Bundesen, C., Shibuya, H., & Larsen, A. (1985). Visual selection from multielement displays: A model for partial report. *Attention and Performance Vol. XI*, 631–649.

Bundesen, Claus, Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review*, *112*(2), 291–328.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *The Journal of Neuroscience*, *21*, 300–304.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18751–18756.

Callejas, A., Lupiáñez, J., & Tudela, P. (2004). The three attentional networks: On their independence and interactions. *Brain & Cognition*, *54*(3), 225–227. doi:10.1016/j.bandc.2004.02.012

Calvert, G. A. (2001). Crossmodal processing in human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110–1123.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649–657.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology*, *98*, 191–205.

Cappe, C., Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways. *Hearing Research*, *258*, 28–36.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: Timing, topography, directionality, and sources. *The Journal of Neuroscience*, *30*(38), 12572–12580.

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *The Journal of Neuroscience*, *31*(25), 9315–9322.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345–347.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experiment Psychology: Human Perception and Performance*, *21*(1), 109–127.

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1993). Ventral intraparietal area of the macaque: Anatomic location and visual response. *Journal of Neurophysiology*, *69*(3), 902–914.

Corbetta, M., Kincade, J. M., Ollinger, J. . M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*(3), 292–7.

Corbetta, M., Kincade, J. M., & Shulman, G. L. (2002). Neural systems for visual orienting and their relationships to spatial working memory. *Journal of Cognitive Neuroscience*, *14*(3), 508–523.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience*, *18*(18), 7426–7435.

Coull, J. T., Nobre, A. C., & Frith, C. D. (2001). The noradrenergic alpha2 agonist clonidine modulates behavioural and neuroanatomical correlates of human attentional orienting and alerting. *Cerebral Cortex*, *11*(1), 73–84.

Dalton, P., & Lavie, N. (2004). Auditory attentional capture: Effects of singleton distractor sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 180–193.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Di Lollo, V., Kawahara, J., Shahab Ghorashi, S. M., & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological Research*, *69*(3), 191–200.

Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *The Journal of Neuroscience*, *25*(36), 8259–8266.

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, *3*(3), 277–283.

Downing, C. J., & Pinker, S. (1985). *The spatial structure of visual attention*. Cambridge, Massachusetts: MIT Press.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgements. *Neuron*, *57*, 11–23.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517.

Duncan, J. (2006). EPS Mid-Career Award 2004. Brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology*, *59*(1), 2–27.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 578–588.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, *7*(2), 255–261.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*, 808–810.

Eardley, A. F., & Van Velzen, J. (2011). Event-related potential evidence for the use of external coordinates in the preparation of tactile attention by the early blind. *European Journal of Neuroscience*, *33*(10), 1897–1907.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161–77.

Eimer, M. (1994). "Sensory gating" as a mechanism for visuospatial orienting: electrophysiological evidence from trial-by-trial cuing experiments. *Perception & Psychophysics*, *55*(6), 667–675.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234.

Eimer, M. (1999). Can attention be directed to opposite locations in different modalities? An ERP study. *Clinical Neurophysiology*, *110*, 1252–1259.

Eimer, M., & Driver, J. (2000). An event-related brain potential study of cross-modal links in spatial attention between vision and touch. *Psychophysiology*, *37*(5), 697–705.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*, 15–31.

Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*, 1423–1433.

Eimer, M., Kiss, M., & Nicholas, S. (2011). What top-down task sets do for us: An ERP study on the benefits of advance preparation in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1758–1766.

Eimer, M., Kiss, M., Press, C., & Sauter, D. (2009). The roles of feature-specific task set and bottom-up salience in attentional capture: An ERP study. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1316–1328.

Eimer, M., & Schroeger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology*, *35*, 313–327.

Eimer, M., Van Velzen, J., & Driver, J. (2002). Cross-modal interactions between audition, touch, and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations. *Journal of Cognitive Neuroscience*, *14*, 254–271.

Eimer, Martin, & Kiss, M. (2010). Top-down search strategies determine attentional capture in visual search: Behavioral and electrophysiological evidence. *Attention, Perception, & Psychophysics*, *72*(4), 951–962.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240.

Eriksen, C. W., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experiment Psychology: Human Perception and Performance*, *11*(5), 583–597.

Eriksen, Charles W, & Murphy, T. D. (1987). Movement of attentional focus across the visual. *Perception & Psychophysics*, *42*(3), 299–305.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.

Evans, E. F., & Whitfield, I. C. (1964). Classification of unit responses in the auditory cortex of the unanaesthisized and unrestrained cat. *Journal of Physiology*, *171*, 476–493.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 1–12.

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, *29*, 1247–1257.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience*, *22*(13), 5749–5759.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage*, *26*, 471 – 479.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347.

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrowt, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia*, *27*, 461–470.

Feldman, J. A. (1985). Connectionist models and parallelism in high level vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 178–200.

Fernandez-Duque, D., & Posner, M. I. (1997). Relating the mechanisms of orienting and alerting. *Neuropsychologia*, *35*, 477–486.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex*, *20*, 109–120.

Fink, G. R., Dolan, R. J., Halligan, P. W., Marshall, J. C., & Frith, C. D. (1997). Space-based and object-based visual attention: shared and specific neural domains. *Brain*, *120*, 2013–2028.

Fockert, J. De, Rees, G., Frith, C., & Lavie, N. (2004). Neural correlates of sttentional capture in visual search. *Journal of Cognitive Neuroscience*, *16*(5), 751–759.

Folk, C. L., Ester, E. F., & Troemel, K. (2009). How to keep attention from straying: Get engaged! *Psychonomic Bulletin & Review*, *16*(1), 127–132.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 847–858.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002a). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*, *14*, 20–30.

Fort, A., Delpuech, C., Pernier, J., & Giard, M.-H. (2002b). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex*, *12*, 1031–1039.

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimension-weighting " account. *Perception & Psychophysics*, *58*(1), 88–101.

Fournier, L. R., & Eriksen, C. W. (1990). Coactivation in the perception of redundant targets. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 538–550.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, *10*, 77–83.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport*, *16*, 419–423.

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., Ritter, W., & Murray, M. M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: An fMRI study. *Journal of Neurophysiology*, *88*, 540–543.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*, 332–343.

Fries, W. (1985). Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase. *The Journal of Comparative Neurology*, *230*(1), 55–76.

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Science*, *273*, 865–874.

Funes, M. J., Lupiáñez, J., & Milliken, B. (2007). Separate mechanisms recruited by exogenous and endogenous spatial cues: Evidence from a spatial Stroop paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 348–362.

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*, 347–351.

Ghazanfar, A. a, & Chandrasekaran, C. F. (2007). Paving the way forward: Integrating the senses through phase-resetting of cortical oscillations. *Neuron*, *53*(2), 162–4.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Cerebral Cortex*, *25*(20), 5004 –5012.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory□? *Trends in Cognitive Sciences*, *10*(6).

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–90.

Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 699–706.

Gillmeister, H., & Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Research*, *1160*, 58–68.

Girelli, M., & Luck, S. J. (1997). Are the same attentional mechanisms used to detect visual search targets defined by color, orientation, and motion? *Journal of Cognitive Neuroscience*, *9*(2), 238–253.

Gondan, M., Goetze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics*, *72*, 1692–1709.

Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (2005). Multisensory processing in the redundant-target effect: A behavioral and event-related potential study. *Perception & Psychophysics*, *67*, 713–726.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25.

Gramann, K., Toellner, T., Krummenacher, J., Eimer, M., & Müller, H. J. (2007). Brain electrical correlates of dimensional weighting: An ERP study. *Psychophysiology*, *44*, 277–292.

Graziano, M. S. A. (2001). Is reaching eye-centered, body-centered, hand-centered, or a combination? *Reviews in the Neurosciences*, *12*, 175–186.

Graziano, M. S. A., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science*, *266*, 1054–1057.

Griffin, I. C., Miniussi, C., & Nobre, A. C. (2001). Orienting attention in time. *Frontiers in Bioscience*, *12*, 660–671.

Grubert, A., Krummenacher, J., & Eimer, M. (2011). Redundancy gains in pop-out visual search are determined by top-down task set: Behavioral and electrophysiological evidence. *Journal of Vision*, *11*, 1–10.

Grunewald, A., Linden, J. F., & Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area I . Effects of training responses to auditory stimuli. *Journal of Neurophysiology*, *82*, 330–342.

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2001). Audiotactile interactions in roughness perception. *Experimental Brain Research*, *146*, 161–171.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *The Journal of Comparative Neurology*, *502*, 924–952.

Hackley, S. A. (2009). The speeding of voluntary reaction by a warning signal. *Psychophysiology*, *46*, 225–233.

Hackley, S. A., Woldorff, M. G., & Hillyard, S. A. (1990). Cross-modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials. *Psychophysiology*, *27*, 195–208.

Harter, M. R., Miller, S. L., Price, N. J., LaLonde, M. E., & Keyes, A. L. (1989). Neural processes involved in directing attention. *Journal of Cognitive Neuroscience*, *1*(3), 223–237.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *The Journal of Neuroscience*, *14*(11), 6336–6353.

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *The Journal of Neuroscience*, *27*(30), 7881–7887.

Helbig, H. B., & Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of Vision*, *8*, 1–16.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*, 289–293.

Hickey, C., Di Lollo, V., & McDonald, J. J. (2008). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience*, *21*, 760–775.

Hickey, Clayton, McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, *18*(4), 604–613.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 781–787.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society B*, *353*, 1257–1270.

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals□: when spatial correspondence matters. *Experimental Brain Research*, *195*, 261–272.

Hocking, J., & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex*, *18*, 2439–2449.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Brain*, *45*, 3340–3345.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology*, *15*(18), 762–764.

Hopf, J.-M., Boelmans, K., Schoenfeld, A. M., Heinze, H., & Luck, S. J. (2002). How does attention attenuate target-distractor interference in vision? Evidence from magnetoencephalographic recordings. *Cognitive Brain Research*, *15*, 17–29.

Hopf, J.-M., Luck, S. J., Girelli, M., Mangun, G. R., Scheich, H., & Heinze, H.-J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, *10*(12), 1233–1241.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 131–153.

Iordanescu, L., Grabowecky, M., Franconeri, S. L., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception & Psychophysics*, *72*(7), 1736–1741.

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, *15*(3), 548–554.

Irons, J. L., & Remington, R. W. (2013). Can attentional control settings be maintained for two color-location conjunctions? Evidence from an RSVP task. *Attention, Perception, & Psychophysics*, 1–14. doi:10.3758/s13414-013-0439-8

James, W. (1890). *The Principles of Psychology*. London: MacMillan.

Jaśkowski, P. (1993). Temporal-order judgment and reaction time to stimuli of different rise times. *Perception*, *22*(8), 963–70.

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology*, *85*, 506–522.

Jones, E. G., & Powell, T. P. S. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, *93*, 793–820.

Kastner, S., Pinsk, M. a, De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure & Function*, *212*, 121–132.

Kennett, S., Eimer, M., Spence, C., & Driver, J. (2001). Tactile-visual links in exogenous spatial attention under different postures: Convergent evidence from psychophysics and ERPs. *Journal of Cognitive Neuroscience*, *13*(4), 462–478.

Kiss, M., & Eimer, M. (2011). Attentional capture by size singletons is determined by top-down search goals. *Psychophysiology*, *48*(6), 784–787.

Kiss, M., Grubert, A., & Eimer, M. (2013). Top-down task sets for combined features: Behavioral and electrophysiological evidence for two stages in attentional object selection. *Attention, Perception & Psychophysics*, *75*(2), 216–228.

Kiss, M., Grubert, A., Petersen, A., & Eimer, M. (2012). Attentional capture by salient distractors during visual search is determined by temporal task demands. *Journal of Cognitive Neuroscience*, *24*(3), 749–59.

Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology*, *45*, 240–249.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*, 372–384.

Kösem, A., & Van Wassenhove, V. (2012). Temporal structure in audiovisual sensory selection. *PLoS ONE*, *7*(7), e40936.

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–292.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, *94*(3), 1904–1911.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579.

Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 1003–1020.

Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(6), 1019–1031.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4), 405–414.

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, *9*, 75–82.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*, 143–148.

Lavie, N., Hirst, A., Fockert, J. De, & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*, 339–354.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197.

Leblanc, E., Prime, D. J., & Jolicoeur, P. (2008). Tracking the location of visuospatial attention in a contingent capture paradigm. *Journal of Cognitive Neuroscience*, *20*(4), 657–671.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(18), 10530–10535.

Lieberman, M. D. (1996). *Speech: A special code*. Cambridge, MA: The MIT Press.

Lien, M.-C., Ruthruff, E., Goodin, Z., & Remington, R. W. (2008). Contingent attentional capture by top-down control settings: Converging evidence from event-related potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 509–530. doi:10.1037/0096-1523.34.3.509

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, *1173*, 102–109.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: Bradford/ MIT Press.

Luck, S. J., & Beach, N. J. (1998). Visual attention and the binding problem: A neurophysiological perspective. *Visual Attention* (pp. 455–475). New York, NY: Oxford University Press.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology*, *33*(1), 64–87.

Luck, S. J., & Hillyard, S. A. (1994a). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291–308.

Luck, S. J., & Hillyard, S. A. (1994b). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 1000–1014.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences*, *28*(5), 264–271.

Macaluso, E., Frith, C. D., & Driver, J. (2001). Multimodal mechanisms of attention related to rates of spatial shifting in vision and touch. *Experimental Brain Research*, *137*, 445–454.

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*, 4–18.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experiment Psychology: Human Perception and Performance*, *17*(4), 1057–1074.

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*, *18*(5), 904–909.

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *10*, 363–401.

Mazza, V., Turatto, M., & Caramazza, A. (2009). Attention selection, distractor suppression and N2pc. *Cortex*, *45*(7), 879–890. doi:10.1016/j.cortex.2008.10.009

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research*, *181*(3), 531–536.

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*, *15*(1), 10–19. doi:10.1162/089892903321107783

McDonald, J. J., Teder-Sälejärvi, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science*, *292*, 1791.

McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science*, *11*(2), 167–171.

McDonald, J. J., Ward, L. M., & Kiehl, K. A. (1999). An event-related brain potential study of inhibition of return. *Perception & Psychophysics*, *61*(7), 1411–1123.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration neurons. I. Temporal factors in superior colliculus. *The Journal of Neuroscience*, *7*(10), 3215–3229.

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, *365*(2), 350–354.

Mesulam, M. M., & Mufson, E. J. (1984). Neural inputs into the nucleus basalis of the substantia innominata (Ch4) in the rhesus monkey. *Brain*, *107*(1), 253–274.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 160–169.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*, 99–115.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*, 57–77.

Molholm, S., Ritter, W., & Javitt, D. C. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, *14*, 452–465.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, *14*, 115–128.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163.

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 1680–1685.

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, *20*, 408–434.

Müller, H.J., & Krummenacher, J. (2006). Locus of dimension weighting: Preattentive or postselective? *Visual Cognition*, *14*, 490–513.

Müller, Hermann J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Perception*, *15*(2), 315–330.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., Charles, E., et al. (2005). Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*, *15*, 963–974. doi:10.1093/cercor/bhh197

Naumer, M. J., Doehrmann, O., Müller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2009). Cortical plasticity of audio-visual object representations. *Cerebral Cortex*, *19*(7), 1641–1653.

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, *12*, 37–42.

Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, *72*(6), 1654–1665.

Nobre, A. C., Sebestyen, G. N., & Miniussi, C. (2000). The dynamics of shifting visuospatial attention revealed by event-related brain potentials. *Neuropsychologia*, *38*, 964–974.

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, *1220*, 157–163.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: correlates of "popout" under anesthesia. *Visual Neuroscience*, *16*(1), 15–34.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*(11), 1203–1209.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. G. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, *65*, 123–132.

Olivers, C. N. L. (2011). Long-term visual associations affect attentional guidance. *Acta Psychologica*, *137*(2), 243–247.

Olivers, C. N. L., & Eimer, M. (2011). On the difference between working memory and attentional set. *Neuropsychologia*, *49*(6), 1553–1558.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: when it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334.

Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, *1242*, 191–199.

Olivers, C. N. L., Van der Stigchel, S., & Hulleman, J. (2007). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research*, *71*(2), 126–139.

Olivers, C. N.L. (2007). The time course of attention: it is better than we though. *Current Directions in Psychological Science*, *16*(1), 11–15.

Pavani, F., Spence, C., & Driver, J. (2000). Visual capture of touch: Out-of-the-body experience with rubber gloves. *Psychological Science*, *11*, 353–359.

Pearson, R. C. A., Brodal, P., Gatter, K. C., & Powell, T. P. S. (1982). The organization of the connections between the cortex and the claustrum in the monkey. *Brain Research*, *234*(2), 435–441.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bonwhuis (Eds.), *Attention and Performance X: Control of Language Processes* (pp. 551–556). Hillsdale, NJ: Erlbaum.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, *58*, 1–23.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, *129*(5), 643–73.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experiment Psychology: Human Perception and Performance*, *18*(3), 849–860.

Rees, G., Frith, C., & Lavie, N. (2001). Processing of irrelevant visual motion during performance of an auditory attention task. *Neuropsychologia*, *39*(9), 937–49.

Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, *37*(5), 8538–63.

Robinson, D. L., Bowman, E. M., & Kertzman, C. (1995). Covert orienting of attention in macaques. II. Contributions of parietal cortex. *Journal of Neurophysiology*, *74*(2), 698–712.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, *50*, 19–26.

Russo, F. Di, Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2001). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *111*(15), 95–111.

Sanabria, D., Soto-Faraco, S., Chan, J., & Spence, C. (2005). Intramodal perceptual grouping modulates multisensory integration: Evidence from the crossmodal dynamic capture task. *Neuroscience Letters*, *377*, 59–64.

Sanabria, D., Soto-Faraco, S., & Spence, C. (2004). Exploring the role of visual perceptual grouping on the audiovisual integration of motion. *Neuroreport*, *15*(18), 2745–2749.

Santangelo, V., Finoia, P., Raffone, A., Belardinelli, M. O., & Spence, C. (2008). Perceptual load affects exogenous spatial orienting while working memory load does not. *Experimental Brain Research.*, *184*(3), 371–82.

Santangelo, V., & Spence, C. (2007a). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1311–1321.

Santangelo, V., & Spence, C. (2007b). Assessing the automaticity of the exogenous orienting of tactile attention. *Perception*, *36*(10), 1497–1505.

Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and Cognition*, *17*, 989–1015.

Santangelo, V., Van der Lubbe, R. H. J., Olivetti Belardinelli, M., & Postma, A. (2006). Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: A comparison of reflexive orienting mechanisms. *Experimental Brain Research*, *173*(1), 40–48.

Santangelo, V., Van der Lubbe, R. H. J., Olivetti Belardinelli, M., & Postma, A. (2008). Multisensory integration affects ERP components elicited by exogenous cues. *Experimental Brain Research*, *185*(2), 269–277.

Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, & Psychophysics*, *72*(6), 1455–1470.

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, "unisensory" processing. *Current Opinion in Neurobiology*, *15*(4), 454–458.

Seiss, E., Kiss, M., & Eimer, M. (2009). Does focused endogenous attention prevent attentional capture in pop-out visual search? *Psychophysiology*, *46*, 703–717.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science*, *16*(2), 114–122.

Shams, L., Kamitani, C. A. Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Cognitive Neuroscience and Neuropsychology*, *12*(17), 3849–3852.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147–152.

Shapiro, K. L., Arnell, K. M., & Raymond, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296.

Shapiro, K. L., & Raymond, J. E. (1994). Temporal allocation of visual attention. Inhibition or interference? In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory Processes in Attention, Memory, and Language* (pp. 151–188). San Diego, CA: Academic Press.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, *145*, 1328–1330.

Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual prior entry. *Psychological Science*, *12*(3), 205–212.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, *12*(1), 7–10.

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology*, *17*(19), 1680–1685.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research*, *14*, 139–146.

Spector, F., & Maurer, D. (2009). Synesthesia: A new approach to understanding the development of perception. *Developmental Psychology*, *45*(1), 175–189.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experiment Psychology: Human Perception and Performance*, *22*(4), 1005–1030.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*(1), 1–22.

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *The Journal of Neuroscience*, *25*(28), 6499–6508.

Stein, B. E, & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–267.

Stein, B. E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*, *123*, 124–135.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*(6), 497–506.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stein, B. E., Meredith, M. A., Huneycutt, W. A., & McDade, L. (1988). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, *1*, 12–24.

Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, *198*, 113–126.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2009). The neural basis of multisensory integration in the midbrain: Its organization and maturation. *Hearing Research*, *258*(1-2), 4–15.

Stricanne, B., Andersen, R. A., & Mazzoni, P. (1996). Eye-centered , head-centered , and intermed iate cod ing of remembered sound locations in srea LIP. *Journal of Neurophysiology*, *76*(3), 2071–2076.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, *49*(6), 905–916.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*, 679.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*, 400–410.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*, 1098–1114.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.

Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(21), 8239–8244.

Teder-Sälejärvi, A. W., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*, 106–114.

Teder-Sälejärvi, W. A., Münte, T. F., Sperlich, F., & Hillyard, S. A. (1999). Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. *Cognitive Brain Research*, *8*(3), 327–343.

Tellinghuisen, D. J., & Nowak, E. J. (2003). The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception & Psychophysics*, *65*(5), 817–828.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, *50*(2), 184–193.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experiment Psychology: Human Perception and Performance*, *20*(4), 799–806.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2), 77–99.

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control visual attention. *Attention and Performance XVIII*, 105–124.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 459–78.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *Journal of Comparative Neurology*, *248*, 190–222.

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE*, *5*(5), e10664.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008a). Pip and Pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053–1065. doi:10.1037/0096-1523.34.5.1053

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008b). Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision*, *8*(5), 1–10.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile-visual synchrony increases visual saliency. *Neuroscience Letters*, *450*(1), 60–64.

Van der Burg, E., Olivers, C. N. L., & Theeuwes, J. (2012). The attentional window modulates capture by audiovisual events. (J. J. Geng, Ed.)*PLoS ONE*, *7*(7), e39137.

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, *55*, 1208–1218.

Van Velzen, J., Eardley, A. F., Forster, B., & Eimer, M. (2006). Shifts of attention in the early blind: An ERP study of attentional control processes in the absence of visual spatial information. *Neuropsychologia*, *44*(12), 2533–2546.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*, 744–56.

Vogel, E. . K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.

Vroomen, J., Bertelson, P., & De Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica*, *108*, 21–33.

Vroomen, J., Bertelson, P., & De Gelder, B. (2001b). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–659.

Vroomen, J., & De Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1583–1590.

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1063–1071.

Wallace, M. T. (2004). The development of multisensory processes. *Cognitive Processing*, *5*(2), 69–83.

Wallace, M. T., Carriere, B. N., Perrault, T. J., Vaughan, J. W., & Stein, B. E. (2006). The development of cortical multisensory integration. *The Journal of Neuroscience*, *26*(46), 11844–11849.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology*, *80*, 1006–1010.

Wallace, M. T., Ramachandran, R., & Stein, B. E. (2004). A revised view of sensory cortical parcellation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(7), 2167–2172.

Warden, M. R., & Miller, E. K. (2007). The representation of multiple objects in prefrontal neuronal delay activity. *Cerebral Cortex*, *17*, 441–450.

Watson, D. G., Humphreys, G. W., & Olivers, C. N. L. (2003). Visual marking: Using time in visual selection. *Trends in Cognitive Sciences*, *7*(4), 180–186.

Weissman, D. H., Warner, L. M., & Woldorff, M. G. (2004). The neural mechanisms for minimizing cross-modal distraction. *The Journal of Neuroscience*, *24*(48), 10941–10949.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience*, *30*(7), 2662–2675.

Wolfe, J. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–67). Hove: Psychology Press.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. doi:10.3758/BF03200774

Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99–120). New York, NY: Oxford University Press.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experiment Psychology: Human Perception and Performance*, *15*(3), 419–433.

Wolfe, J. M., Chun, M. M., & Friedman-Hill, S. R. (1995). Making use of texton gradients: Visual search and perceptual grouping exploit the same parallel processes in different ways. In T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early Vision and Beyond* (pp. 189–197). Cambridge, MA: The MIT Press.

Woodman, G. F., Arita, J. T., & Luck, S. J. (2009). A cuing study of the N2pc component: An index of attentional deployment to objects rather than spatial locations. *Vision Research, 1297*, 101–111. doi:10.1016/j.brainres.2009.08.011

Woods, D. L., Alho, K., & Algazi, A. (1992). Intermodal selective attention: I. Effects on event-related potentials to lateralized auditory and visual stimuli. *Electroencephalography & Clinical Neurophysiology*, *82*, 341–355.

Yamaguchi, S., Tsuchiya, H., & Kobayashi, S. (1998). Visuospatial attention shift and motor responses in cerebellar disorders. *Journal of Cognitive Neuroscience*, *10*(1), 95–107.

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 661–676.

Appendices

Appendix 1

Experiment 1

Reaction times:

- Main effect of cue type, $F(1,19) = .46$, $p = .51$, $\eta_p^2 = .02$,
- Main effect of tone presence, $F(1,19) = 13.84$, $p < .001$, $\eta_p^2 = .42$,
- Main effect of spatial cueing, $F(1,19) = 172.86$, $p < .001$, $\eta_p^2 = .9$,
- Interaction of cue type and tone presence, $F(1,19) = .12$, $p = .73$, $\eta_p^2 = .01$,
- Interaction of cue type and spatial cueing, $F(1,19) = 2.86$, $p = .11$, $\eta_p^2 = .13$,
- Interaction of tone presence and spatial cueing, $F(1,19) = 4.71$, $p < .05$, $\eta_p^2 = .2$,
- Interaction of cue type, tone presence and spatial cueing, $F(1,19) = .77$, $p = .39$, $\eta_p^2 = .04$.

Error rates:

- Main effect of cue type, $F(1,19) = .03$, $p = .88$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,19) = 5.49$, $p < .05$, $\eta_p^2 = .22$,
- Main effect of spatial cueing, $F(1,19) = 57.62$, $p < .001$, $\eta_p^2 = .75$,
- Interaction cue type and tone presence, $F(1,19) = .71$, $p = .41$, $\eta_p^2 = .04$,
- Interaction cue type and spatial cueing, $F(1,19) = .02$, $p = .89$, $\eta_p^2 = .01$,
- Interaction tone presence and spatial cueing, $F(1,19) = 4.79$, $p < .05$, $\eta_p^2 = .2$,
- Interaction cue type, tone presence and spatial cueing, $F(1,19) = .03$, $p = .88$, $\eta_p^2 = .01$.

Appendix 2

Results of Experiment 2

Reaction times:

- Main effect of cue type, $F(1,11) = .08$, $p = .79$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,11) = 1.79$, $p = .21$, $\eta_p^2 = .14$,
- Main effect of spatial cueing, $F(1,11) = 31.03$, $p < .001$, $\eta_p^2 = .74$,
- Interaction of cue type and tone presence, $F(1,11) = .44$, $p = .52$, $\eta_p^2 = .04$,
- Interaction of cue type and spatial cueing, $F(1,11) = .72$, $p = .41$, $\eta_p^2 = .06$,
- Interaction of tone presence and spatial cueing, $F(1,11) = .59$, $p = .46$, $\eta_p^2 = .05$,
- Interaction of cue type, tone presence and spatial cueing, $F(1,11) = .01$, $p = .92$, $\eta_p^2 = .01$.

Error rates:

- Main effect of cue type, $F(1,11) = .03$, $p = .86$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,11) = 5.19$, $p < .05$, $\eta_p^2 = .2$,
- Main effect of spatial cueing, $F(1,11) = 12.6$, $p < .01$, $\eta_p^2 = .38$,
- Interaction cue type and tone presence, $F(1,11) = .02$, $p = .91$, $\eta_p^2 = .01$,
- Interaction cue type and spatial cueing, $F(1,11) = 1.51$, $p = .23$, $\eta_p^2 = .07$,
- Interaction tone presence and spatial cueing, $F(1,11) = 1.71$, $p = .11$, $\eta_p^2 = .08$,
- Interaction cue type, tone presence and spatial cueing, $F(1,11) = 1.79$, $p = .2$, $\eta_p^2 = .08$.

Appendix 3

Results of Experiment 3

Reaction times:

- Main effect of cue type, $F(1,21) = 1.58$, $p = .22$, $\eta_p^2 = .07$,
- Main effect of tone presence, $F(1,21) = 16.31$, $p < .001$, $\eta_p^2 = .44$,
- Main effect of spatial cueing, $F(1,21) = 18.31$, $p < .001$, $\eta_p^2 = .47$,
- Interaction of cue type and tone presence, $F(1,21) = .23$, $p = .63$, $\eta_p^2 = .01$,
- Interaction of cue type and spatial cueing, $F(1,21) = 20.95$, $p < .001$, $\eta_p^2 = .5$,
- Interaction of tone presence and spatial cueing, $F(1,21) = 2.47$, $p = .13$, $\eta_p^2 = .11$,
- Interaction of cue type, tone presence and spatial cueing, $F(1,21) = .06$, $p = .81$, $\eta_p^2 = .01$.

Error rates:

- Main effect of cue type, $F(1,21) = .33$, $p = .57$, $\eta_p^2 = .02$,
- Main effect of tone presence, $F(1,21) = 3.75$, $p = .07$, $\eta_p^2 = .15$,
- Main effect of spatial cueing, $F(1,21) = 5.85$, $p < .05$, $\eta_p^2 = .22$,
- Interaction cue type and tone presence, $F(1,21) = .02$, $p = .89$, $\eta_p^2 = .01$,
- Interaction cue type and spatial cueing, $F(1,21) = 5.2$, $p < .05$, $\eta_p^2 = .2$,
- Interaction tone presence and spatial cueing, $F(1,21) = .03$, $p = .86$, $\eta_p^2 = .01$,
- Interaction cue type, tone presence and spatial cueing, $F(1,21) = .04$, $p = .83$, $\eta_p^2 = .01$.

Appendix 4

Results of Experiment 4

Reaction times:

- Main effect of cue type, $F(1,21) = .01$, $p = .99$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,21) = 8.33$, $p < .01$, $\eta_p^2 = .28$,
- Main effect of spatial cueing, $F(1,21) = 21.07$, $p < .001$, $\eta_p^2 = .5$,
- Interaction of cue type and tone presence, $F(1,21) = 3.09$, $p = .09$, $\eta_p^2 = .13$,
- Interaction of cue type and spatial cueing, $F(1,21) = 1.46$, $p = .24$, $\eta_p^2 = .07$,
- Interaction of tone presence and spatial cueing, $F(1,21) = .58$, $p = .46$, $\eta_p^2 = .03$,
- Interaction of cue type, tone presence and spatial cueing, $F(1,21) = 2.28$, $p = .15$, $\eta_p^2 = .1$.

Error rates:

- Main effect of cue type, $F(1,21) = .03$, $p = .86$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,21) = 5.19$, $p < .05$, $\eta_p^2 = .2$,
- Main effect of spatial cueing, $F(1,21) = 12.6$, $p < .01$, $\eta_p^2 = .38$,
- Interaction cue type and tone presence, $F(1,21) = .02$, $p = .91$, $\eta_p^2 = .01$,
- Interaction cue type and spatial cueing, $F(1,21) = 1.51$, $p = .23$, $\eta_p^2 = .07$,
- Interaction tone presence and spatial cueing, $F(1,21) = 1.71$, $p = .21$, $\eta_p^2 = .08$,
- Interaction cue type, tone presence and spatial cueing, $F(1,21) = 1.79$, $p = .2$, $\eta_p^2 = .08$.

Combined RTs analysis of Experiments 3 & 4

- Interaction of cue type and visual selectivity, $F(1,42) = .92$, $p = .34$, $\eta_p^2 = .02$,
- Interaction of tone presence and visual selectivity, $F(1,42) = 2.23$, $p = .14$, $\eta_p^2 = .05$,
- Interaction of spatial cueing and visual selectivity, $F(1,42) = .55$, $p = .46$, $\eta_p^2 = .01$,
- Interaction of cue type, tone presence and visual selectivity, $F(1,42) = 2.6$, $p = .12$, $\eta_p^2 = .06$,
- Interaction of cue type, spatial cueing and visual selectivity, $F(1,42) = 5.$, $p < .05$, $\eta_p^2 = .11$,

- Interaction of tone presence, spatial cueing and visual selectivity, $F(1,42) = .29$, $p = .59$, $\eta_p^2 = .01$,
- Interaction of cue type, tone presence, spatial cueing and visual selectivity, $F(1,42) = 1.58$, $p = .22$, $\eta_p^2 = .04$.

Combined RTs analysis of Experiments 1 & 4

- Interaction of cue type and relative salience, $F(1,40) = .16$, $p = .69$, $\eta_p^2 = .01$,
- Interaction of tone presence and relative salience, $F(1,40) = .17$, $p = .68$, $\eta_p^2 = .01$,
- Interaction of spatial cueing and relative salience, $F(1,40) = 34.98$, $p < .001$, $\eta_p^2 = .47$,
- Interaction of cue type, tone presence and relative salience, $F(1,40) = 2.16$, $p = .15$, $\eta_p^2 = .05$,
- Interaction of cue type, spatial cueing and relative salience: $F(1,40) = .19$, $p = .67$, $\eta_p^2 = .01$,
- Interaction of tone presence, spatial cueing and relative salience, $F(1,40) = 4.11$, $p < .05$, $\eta_p^2 = .09$,
- Interaction of cue type, tone presence, spatial cueing and relative salience, $F(1,40) = 2.91$, $p = .1$, $\eta_p^2 = .07$.

Appendix 5

Results of Experiment 5

Reaction times:

- Main effect of cue type, $F(1,21) = 2.7$, $p = .12$, $\eta_p^2 = .11$,
- Main effect of tone presence, $F(1,21) = 32.83$, $p < .001$, $\eta_p^2 = .61$,
- Main effect of spatial cueing, $F(1,21) = 47.49$, $p < .001$, $\eta_p^2 = .69$,
- Interaction of cue type and tone presence, $F(1,21) = .53$, $p = .47$, $\eta_p^2 = .03$,
- Interaction of cue type and spatial cueing, $F(1,21) = 39.37$, $p < .001$, $\eta_p^2 = .65$,
- Interaction of tone presence and spatial cueing, $F(1,21) = 4.5$, $p < .05$, $\eta_p^2 = .18$,
- Interaction of cue type, tone presence and spatial cueing, $F(1,21) = .04$, $p = .85$, $\eta_p^2 = .01$.

Error rates:

- Main effect of cue type, $F(1,21) = 1.82$, $p = .19$, $\eta_p^2 = .08$,
- Main effect of tone presence, $F(1,21) = 10.66$, $p < .01$, $\eta_p^2 = .34$,
- Main effect of spatial cueing, $F(1,21) = 21.83$, $p < .001$, $\eta_p^2 = .51$,
- Interaction cue type and tone presence, $F(1,21) = .13$, $p = .72$, $\eta_p^2 = .01$,
- Interaction cue type and spatial cueing, $F(1,21) = 8.97$, $p < .01$, $\eta_p^2 = .3$,
- Interaction tone presence and spatial cueing, $F(1,21) = 4.32$, $p = .5$, $\eta_p^2 = .17$,
- Interaction cue type, tone presence and spatial cueing, $F(1,21) = .33$, $p = .57$, $\eta_p^2 = .02$.

Combined RTs analysis of Experiments 4 & 5

- Interaction of cue type and tone intensity, $F(1,42) = .32$, $p = .58$, $\eta_p^2 = .01$,
- Interaction of tone presence and tone intensity, $F(1,42) = .5$, $p = .49$, $\eta_p^2 = .01$,
- Interaction of spatial cueing and tone intensity, $F(1,42) = .58$, $p = .45$, $\eta_p^2 = .01$,
- Interaction of cue type, tone presence and tone intensity, $F(1,42) = .02$, $p = .89$, $\eta_p^2 = .01$,
- Interaction of cue type, spatial cueing and tone intensity: $F(1,42) = 1.81$, $p = .19$, $\eta_p^2 = .04$,

- Interaction of tone presence, spatial cueing and tone intensity, $F(1,42) = 8.5$, $p = .006$, $\eta_p^2 = .17$,
- Interaction of cue type, tone presence, spatial cueing and tone intensity, $F(1,42) = .26$, $p = .61$, $\eta_p^2 = .01$.

Appendix 6

Results of Experiment 6

Heterogeneous-cue blocks

Reaction times:

- Main effect of cue type, $F(1,15) = 1.7$, $p = .2$, $\eta_p^2 = .11$,
- Main effect of tone presence, $F(1,15) = 22.24$, $p < .001$, $\eta_p^2 = .59$,
- Main effect of spatial cueing, $F(1,15) = 50.13$, $p < .001$, $\eta_p^2 = .77$,
- Interaction of cue type and tone presence, $F(1,15) = 3.66$, $p = .08$, $\eta_p^2 = .2$,
- Interaction of cue type and spatial cueing, $F(1,15) = 8.82$, $p < .01$, $\eta_p^2 = .37$,
- Interaction of tone presence and spatial cueing, $F(1,15) = .96$, $p = .17$, $\eta_p^2 = .06$,
- Interaction of cue type, tone presence and spatial cueing, $F(1,15) = 1.2$, $p = .28$, $\eta_p^2 = .08$.

Error rates:

- Main effect of cue type, $F(1,15) = .05$, $p = .82$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,15) = 2.82$, $p = .11$, $\eta_p^2 = .16$
- Main effect of spatial cueing, $F(1,15) = 9.44$, $p < .01$, $\eta_p^2 = .39$,
- Interaction cue type and tone presence, $F(1,15) = .62$, $p = .44$, $\eta_p^2 = .04$,
- Interaction cue type and spatial cueing, $F(1,15) = 10.06$, $p < .01$, $\eta_p^2 = .4$,
- Interaction tone presence and spatial cueing, $F(1,15) = 1.58$, $p = .23$, $\eta_p^2 = .1$,
- Interaction cue type, tone presence and spatial cueing, $F(1,15) = .85$, $p = .37$, $\eta_p^2 = .05$.

N2pc results:

- Main effect of cue type, $F(1,15) = .05$, $p = .83$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,15) = 4.88$, $p < .05$, $\eta_p^2 = .25$,
- Main effect of contralaterality, $F(1,15) = .42$, $p = .53$, $\eta_p^2 = .03$,

- Interaction of cue type and tone presence, $F(1,15) = 1.43$, $p = .25$, $\eta_p^2 = .09$,
- Interaction of cue type and contralaterality, $F(1,15) = 1.43$, $p = .12$, $\eta_p^2 = .09$,
- Interaction of tone presence and contralaterality, $F(1,15) = 1.6$, $p = .22$, $\eta_p^2 = .1$,
- Interaction of cue type, tone presence and contralaterality, $F(1,15) = 2.23$, $p = .16$, $\eta_p^2 = .13$.

Homogeneous-cue blocks

Reaction times:

- Main effect of cue type, $F(1,15) = 15.4$, $p < .001$, $\eta_p^2 = .51$,
- Main effect of tone presence, $F(1,15) = 11.8$, $p < .01$, $\eta_p^2 = .44$,
- Main effect of spatial cueing, $F(1,15) = 25.98$, $p < .001$, $\eta_p^2 = .63$,
- Interaction of cue type and tone presence, $F(1,15) = .68$, $p = .42$, $\eta_p^2 = .04$,
- Interaction of cue type and spatial cueing, $F(1,15) = 27.11$, $p < .001$, $\eta_p^2 = .64$,
- Interaction of tone presence and spatial cueing, $F(1,15) = .34$, $p = .57$, $\eta_p^2 = .02$,
- Interaction of cue type, tone presence and spatial cueing, $F(1,15) = .1$, $p = .76$, $\eta_p^2 = .01$.

Error rates:

- Main effect of cue type, $F(1,15) = .01$, $p = .96$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,15) = 1.91$, $p = .19$, $\eta_p^2 = .11$,
- Main effect of spatial cueing, $F(1,15) = 3.1$, $p < .05$, $\eta_p^2 = .16$,
- Interaction cue type and tone presence, $F(1,15) = .16$, $p = .7$, $\eta_p^2 = .01$,
- Interaction cue type and spatial cueing, $F(1,15) = 6.61$, $p < .05$, $\eta_p^2 = .29$,
- Interaction tone presence and spatial cueing, $F(1,15) = .77$, $p = .39$, $\eta_p^2 = .05$,
- Interaction cue type, tone presence and spatial cueing, $F(1,15) = .02$, $p = .9$, $\eta_p^2 = .01$.

N2pc results:

- Main effect of cue type, $F(1,15) = .14$, $p = .72$, $\eta_p^2 = .01$,
- Main effect of tone presence, $F(1,15) = 16.05$, $p < .001$, $\eta_p^2 = .52$,
- Main effect of contralaterality, $F(1,15) = 14.85$, $p < .01$, $\eta_p^2 = .5$,
- Interaction of cue type and tone presence, $F(1,15) = .01$, $p = .97$, $\eta_p^2 = .01$,
- Interaction of cue type and contralaterality, $F(1,15) = 7.65$, $p < .05$, $\eta_p^2 = .34$,
- Interaction of tone presence and contralaterality, $F(1,15) = 3.23$, $p < .05$, $\eta_p^2 = .18$,
- Interaction of cue type, tone presence and contralaterality, $F(1,15) = .52$, $p = .48$, $\eta_p^2 = .03$.

Across-block comparisons

Reaction times:

- Main effect of relative salience, $F(1,15) = 3.26$, $p = .09$, $\eta_p^2 = .18$,
- Interaction of cue type and relative salience , $F(1,15) = 2.43$, $p = .14$, $\eta_p^2 = .14$,
- Interaction of tone presence and relative salience, $F(1,15) = 5.27$, $p < .05$, $\eta_p^2 = .26$,
- Interaction of spatial cueing and relative salience , $F(1,15) = .82$, $p = .38$, $\eta_p^2 = .05$,
- Interaction of cue type, tone presence and relative salience, $F(1,15) = .42$, $p = .53$, $\eta_p^2 = .03$,
- Interaction of cue type, spatial cueing and relative salience, $F(1,15) = 4.36$, $p = .3$, $\eta_p^2 = .07$.
- Interaction of tone presence, spatial cueing and relative salience, $F(1,15) = 1.14$, $p = .054$, $\eta_p^2 = .23$.
- Interaction of cue type, tone presence, spatial cueing and relative salience, $F(1,15) = .73$, $p = .41$, $\eta_p^2 = .05$.

Error rates:

- Main effect of relative salience, $F(1,15) = 8.37$, $p < .05$, $\eta_p^2 = .36$,
- Interaction of cue type and relative salience , $F(1,15) = .43$, $p = .52$, $\eta_p^2 = .03$,
- Interaction of tone presence and relative salience, $F(1,15) = .67$, $p = .43$, $\eta_p^2 = .04$,

- Interaction of spatial cueing and relative salience , $F(1,15) = 3.89$, $p = .067$, $\eta_p^2 = .21$,
- Interaction of cue type, tone presence and relative salience, $F(1,15) = .67$, $p = .43$, $\eta_p^2 = .04$,
- Interaction of cue type, spatial cueing and relative salience, $F(1,15) = 1.53$, $p = .27$, $\eta_p^2 = .09$.
- Interaction of tone presence, spatial cueing and relative salience, $F(1,15) = 2.1$, $p = .17$, $\eta_p^2 = .12$.
- Interaction of cue type, tone presence, spatial cueing and relative salience, $F(1,15) = .56$, $p = .47$, $\eta_p^2 = .04$.

N2pc results:

- Main effect of cue salience, $F(1,17) = .77$, $p = .39$, $\eta_p^2 = .04$,
- Interaction of cue type and cue salience , $F(1,17) = .01$, $p = .97$, $\eta_p^2 = .01$,
- Interaction of tone presence and cue salience, $F(1,17) = .13$, $p = .73$, $\eta_p^2 = .01$,
- Interaction of contralaterality and cue salience, $F(1,17) = 11.04$, $p < .01$, $\eta_p^2 = .39$,
- Interaction of cue type, tone presence and cue salience, $F(1,17) = .17$, $p = .69$, $\eta_p^2 = .01$,
- Interaction of cue type, contralaterality and cue salience, $F(1,17) = 1.92$, $p = .18$, $\eta_p^2 = .1$.
- Interaction of tone presence, contralaterality and cue salience, $F(1,17) = 6.34$, $p < .05$, $\eta_p^2 = .27$.
- Interaction of cue type, tone presence, contralaterality and cue salience, $F(1,17) = 2.21$, $p = .16$, $\eta_p^2 = .12$.

Appendix 7

Results of Experiment 7

N2pc results:

- Main effect of colour-bar type, $F(1,11) = 10.12$, $p < .01$, $\eta_p^2 = .48$,
- Main effect of tone presence, $F(1,11) = 15.07$, $p < .01$, $\eta_p^2 = .58$,
- Main effect of contralaterality, $F(1,11) = 49.12$, $p < .001$, $\eta_p^2 = .82$,
- Interaction of colour-bar type and tone presence, $F(1,11) = .37$, $p = .56$, $\eta_p^2 = .03$,
- Interaction of colour-bar type and contralaterality, $F(1,11) = 19.93$, $p < .001$, $\eta_p^2 = .64$,
- Interaction of tone presence and contralaterality, $F(1,11) = 4.89$, $p < .05$, $\eta_p^2 = .31$,
- Interaction of colour-bar type, tone presence and contralaterality, $F(1,11) = .19$, $p = .67$, $\eta_p^2 = .02$.

Appendix 8

Results of Experiment 8

Reaction times:

- Main effect of task, $F(1.33,14.64) = .43$, $p = .58$, $\eta_p^2 = .04$,
- Main effect of spatial cueing, $F(1,11) = 27.59$, $p < .001$, $\eta_p^2 = .72$,
- Interaction of task and spatial cueing, $F(1.35,14.83) = 9.42$, $p < .01$, $\eta_p^2 = .46$.

Error rates:

- Main effect of task, $F(2,22) = .03$, $p = .97$, $\eta_p^2 = .01$,
- Main effect of spatial cueing, $F(1,11) = 6.26$, $p < .05$, $\eta_p^2 = .36$,
- Interaction of task and spatial cueing, $F(2,22) = .37$, $p = .56$, $\eta_p^2 = .03$.

False Alarms:

- Main effect of task, $F(2,22) = 6.23$, $p < .01$, $\eta_p^2 = .36$,
- Main effect of contralaterality, $F(1,11) = 3,38$, $p = .09$, $\eta_p^2 = .24$,
- Interaction of task and spatial cueing, $F(2,22) = 1.54$, $p = .24$, $\eta_p^2 = .12$.

N2pc results:

- Main effect of task, $F(1.34,14.7) = .63$, $p = .49$, $\eta_p^2 = .05$,
- Main effect of contralaterality, $F(1,11) = 21.72$, $p < .001$, $\eta_p^2 = .66$,
- Interaction of task and contralaterality, $F(2,22) = 3.76$, $p < .05$, $\eta_p^2 = .26$.

Appendix 9

Results of Experiment 9

Reaction times:

- Main effect of task, $F(1,11) = 4.25$, $p = .064$, $\eta_p^2 = .28$,
- Main effect of spatial cueing, $F(1,11) = 64.51$, $p < .001$, $\eta_p^2 = .85$,
- Interaction of task and spatial cueing, $F(1,11) = 4.98$, $p < .05$, $\eta_p^2 = .31$.

Error rates:

- Main effect of task, $F(1,11) = .03$, $p = .87$, $\eta_p^2 = .01$
- Main effect of spatial cueing, $F(1,11) = 4.89$, $p < .05$, $\eta_p^2 = .31$,
- Interaction of task and spatial cueing, $F(1,11) = 2.58$, $p = .14$, $\eta_p^2 = .19$.

False Alarms:

- Main effect of task, $F(1,11) = 1.7$, $p = .22$, $\eta_p^2 = .13$,
- Main effect of spatial cueing, $F(1,11) = .53$, $p = .48$, $\eta_p^2 = .05$,
- Interaction of task and spatial cueing, $F(1,11) = .6$, $p = .48$, $\eta_p^2 = .05$.

N2pc results:

- Main effect of task, $F(1,11) = 1.3$, $p = .28$, $\eta_p^2 = .11$,
- Main effect of spatial cueing, $F(1,11) = 43.37$, $p < .001$, $\eta_p^2 = .8$,
- Interaction of task and spatial cueing, $F(1,11) = 9.68$, $p < .01$, $\eta_p^2 = .47$.

Appendix 10

Results of Experiment 10

Reaction times:

- Main effect of task, $F(2,22) = .64$, $p = .54$, $\eta_p^2 = .06$,
- Main effect of spatial cueing, $F(1,11) = 66.89$, $p < .001$, $\eta_p^2 = .86$,
- Interaction of task and spatial cueing, $F(2,22) = 4.89$, $p < .05$, $\eta_p^2 = .31$.

Error rates:

- Main effect of task, $F(2,22) = .47$, $p = .63$, $\eta_p^2 = .04$,
- Main effect of spatial cueing, $F(1,11) = 2.86$, $p = .12$, $\eta_p^2 = .21$,
- Interaction of task and spatial cueing, $F(2,22) = 1.8$, $p = .19$, $\eta_p^2 = .14$.

False Alarms:

- Main effect of task, $F(2,22) = 6.46$, $p < .01$, $\eta_p^2 = .37$,
- Main effect of spatial cueing, $F(1,11) = 2.96$, $p = .12$, $\eta_p^2 = .21$,
- Interaction of task and spatial cueing, $F(2,22) = .88$, $p = .43$, $\eta_p^2 = .07$.

N2pc results:

- Main effect of task, $F(1.18,12.95) = .06$, $p = .85$, $\eta_p^2 = .01$,
- Main effect of contralaterality, $F(1,11) = 21.93$, $p < .001$, $\eta_p^2 = .67$,
- Interaction of task and contralaterality, $F(2,22) = .16$, $p = .85$, $\eta_p^2 = .02$.

Across-block comparisons

Reaction times:

- Main effect of cue salience, $F(1,22) = 4.44$, $p < .05$, $\eta_p^2 = .17$,
- Interaction of task and cue salience, $F(1.52,33.32) = .96$, $p = .37$, $\eta_p^2 = .04$,
- Interaction of spatial cueing and cue salience, $F(1,22) = 6.49$, $p < .05$, $\eta_p^2 = .23$,

- Interaction of task, spatial cueing and cue salience, $F(2,44) = 4.15$, $p < .05$, $\eta_p^2 = .16$.

N2pc results:

- Main effect of cue salience, $F(1,22) = 2.2$, $p = .15$, $\eta_p^2 = .09$
- Interaction of task and cue salience, $F(1.13,27.03) = .42$, $p = .56$, $\eta_p^2 = .02$,
- Interaction of contralaterality and cue salience, $F(1,22) = 1.84$, $p = .19$, $\eta_p^2 = .08$,
- Interaction of task, contralaterality and cue salience, $F(2,44) = 3.05$, $p < .05$, $\eta_p^2 = .12$.

Appendix 11

Results of Experiment 11

Reaction times:

- Main effect of task, $F(2,22) = 1.49$, $p = .25$, $\eta_p^2 = .12$,
- Main effect of spatial cueing, $F(1,11) = 29.95$, $p < .001$, $\eta_p^2 = .73$,
- Interaction of task and spatial cueing, $F(2,22) = 5.89$, $p < .01$, $\eta_p^2 = .35$.

Error rates:

- Main effect of task, $F(2,22) = .24$, $p = .79$, $\eta_p^2 = .02$,
- Main effect of spatial cueing, $F(1,11) = 11.81$, $p < .01$, $\eta_p^2 = .52$,
- Interaction of task and spatial cueing, $F(2,22) = .26$, $p = .78$, $\eta_p^2 = .03$.

False Alarms:

- Main effect of task, $F(2,22) = 1.27$, $p = .3$, $\eta_p^2 = .1$,
- Main effect of spatial cueing, $F(1,11) = .74$, $p = .41$, $\eta_p^2 = .06$,
- Interaction of task and spatial cueing, $F(2,22) = 3.89$, $p < .05$, $\eta_p^2 = .26$.

N2pc results:

- Main effect of task, $F(1.38,15.13) = .2$, $p = .74$, $\eta_p^2 = .02$,
- Main effect of contralaterality, $F(1,11) = 15.03$, $p < .01$, $\eta_p^2 = .58$,
- Interaction of task and contralaterality, $F(2,22) = 7.4$, $p < .01$, $\eta_p^2 = .4$.